Journal of Applied Psychology

JOHN G. DARLEY, Editor
University of Minnesota

Table of Contents

The Prestige of Industries: R. E. Campbell 1

Some Correlates of the Empathy of the Workers with the Foreman: U. G. Foa 6

The Consistency of Ratings by Peers: D. W. Fiske and J. A. Cox, Jr. 11

Driver Behavior and Scores on the MMPI: P. L. Brown and R. P. Berdie 18

Company Policies and Supervisors' Attitudes toward Supervision: E. S. Stanton 22

A Comparative Study of "Transfer through Generalization" and "Transfer through Identical Elements" in Technical Training: A. I. Siegel, M. Richlin, and P. Federman 27

The Validity of Biographical Data Items for Food Company Salesmen: T. W. Harrell 31

The Effects of Task Complexity and Time Pressure upon Team Productivity: Pauline N. Pepinsky, H. B. Pepinsky, and W. B. Pavlik 34

An Inexpensive Method to Determine the Efficiency of a Television Program: L. Myers, Jr. and E. F. Gardner 39

A New Method for Clustering Jobs: D. B. Orr 44

The Effects of Sleep Loss on Component Movements of Human Motion: S. J. Harris 50

Union and Management Trainees—A Comparative Study of Personality and Occupational Choice: H. M. Bogard 56

American Psychological Association

Volume 44, Number 1 February 1960

Consulting Editors

HAROLD E. BURTT, *Ohio State University*
ALPHONSE CHAPANIS, *Johns Hopkins University*
CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*
LAURENCE S. MCGAUGHRAN, *University of Houston*
QUINN MCNEMAR, *Stanford University*
ALEXANDER MINTZ, *City College of New York*
HAROLD F. ROTHE, *Fairbanks, Morse and Company*
JULIAN B. ROTTER, *Ohio State University*
THOMAS A. RYAN, *Cornell University*
DONALD E. SUPER, *Columbia University*
MILES A. TINKER, *University of Minnesota*
ALFRED C. WELCH, *University of New Mexico*

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: vocational and educational prognosis, diagnosis, and guidance at the secondary and college level; personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research.

Because of the large number of manuscripts submitted, authors should adhere to the rule of "brevity consistent with clarity." The typical manuscript should run to approximately 4,000 words. There is a lag of approximately twelve months between receipt and publication of an article. Authors may request advanced publication if they are prepared to pay the cost of printing the necessary extra pages.

Kenneth E. Clark will become Editor of the *Journal of Applied Psychology*, effective January 1, 1961. Because of the publication lag in this Journal, manscripts from now on should be submitted to the Editor-elect:

Dr. Kenneth E. Clark
Department of Psychology
University of Minnesota
Minneapolis 14, Minnesota

All manuscripts must be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographs or pencil-drawn copies.

Manuscripts must conform to the style requirements described in the *Publication Manual of the American Psychological Association.*

Journal of Applied Psychology

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W.
Washington 6, D. C.

$8.00 per volume $1.50 per issue

ARTHUR C. HOFFMAN, Managing Editor; HELEN ORR, Promotion Manager;
BEATRICE B. HAMLIN, Editorial Assistant

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second class postage paid at Lancaster, Pennsylvania and at additional mailing places.

© 1960 by the American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 44, NO. 1 FEBRUARY 1960

THE PRESTIGE OF INDUSTRIES

ROBERT E. CAMPBELL [1]

Ohio State University

Since the original study by Counts in 1925, there have been numerous occupational prestige investigations. Probably the most ambitious of these was conducted by North and Hatt in 1947. Despite the large number of occupational studies, the related area of industries has been neglected. To the best of the writer's knowledge, the only study of the prestige hierarchy of industries was conducted by Brayfield, Kennedy, and Kendall in 1953. The present study was designed to add to our knowledge of industrial prestige and to provide a frame of reference for future research in this area.

To accomplish this aim, the investigation was concerned with four major questions as follows:

1. What is the nature of the hierarchy of industries according to prestige?
2. Do personal attributes of the respondent such as academic training, father's industry, sex, work experience, military experience, and the industry he hopes to enter influence his evaluation?
3. What are the major determinants of the prestige of an industry?
4. Do respondents from two different geographical regions of the United States agree in rating industries?

METHOD

Fifty-two industries were selected representing each of the 10 major divisions of the Standard Industrial Classification System, 1957. The 10 divisions include the following:

agriculture, forestry, and fisheries
mining
contract construction
manufacturing
transportation, communication, electric, gas and sanitary services
wholesale and retail trade
finance, insurance, and real estate
services
government
nonclassifiable establishments

The following criteria were used to select the 52 industries from the 1500 classified:

(*a*) Twenty-two industries were selected from the earlier study of Brayfield et al. to determine the degree of agreement between the studies. See the starred industries in Table 3.

(*b*) Industries that were thought to be well known by the respondents.

(*c*) Similar industries were chosen to discover the rank difference between them. For example, local bus transportation and interstate bus transportation; public schools and private schools; motels and hotels.

(*d*) Industries which reflected our era. This group included weapons manufacturing, atomic energy plants, labor organizations, and research laboratories.

(*e*) The major military services.

The data were obtained by questionnaire. In addition to rating the 52 industries, the respondent was requested to answer questions about himself (age, sex, college, major, family income, etc.). The rating categories were the same as used by North and Hatt in their study of occupations. See Table 1 for the rating scheme. By using this scheme, the theoretical high that an industry could obtain was 100 and the low was 20.

The respondents were cautioned to express their own opinion and not to be influenced by one person in a particular industry.

To eliminate communication ambiguity and to detect flaws in the instrument, a pilot study was conducted. Form A was administered to 52 Ohio State University engineering students. The form was also reviewed by a total of 20 students and faculty members for improvement. In the actual study, three different forms were used. The forms differed only in the arrangement of the industries. Form A listed the industries in the order designated by the Standard Industrial Classification Manual (1957); Form B had the industries arranged by alphabetical order; and Form C was arranged by random assignment.

[1] The author wishes to thank his advisor, C. L. Shartle, for his help with the preparation of this study.

TABLE 1

RATING CATEGORIES

Rating	Score Value	Statement
1	100	This industry has excellent prestige
2	80	This industry has good prestige
3	60	This industry has average prestige
4	40	This industry has fair prestige
5	20	This industry has poor prestige
X	0	I don't know how to rate this industry

The industries were rated by 360 junior and senior students at the Ohio State University. The students were randomly selected from the College of Arts and Sciences and the College of Commerce and Administration. The Commerce sample was limited to business-oriented students to provide an academic contrast with the Arts and Science students. The students represented more than twenty different acedemic majors (banking, journalism, advertising, fine arts, premedical, accounting, etc.).

TABULATION PROCEDURE

A rank order of the industries was established for each group of respondents based on the mean scores. Thus, the mean scores for the men represented one hierarchy of industries. Rank orders were also computed for women, respondents with work experience, those with no work experience, Commerce students, and Arts and Science students. Rank-order correlations were computed for each group comparison (men vs. women, Commerce vs. Arts and Sciences, etc.).

To determine whether the industry of the father influenced the respondent's ratings, the respondents were grouped by their father's industry. The respondent was placed in one of four industrial groups (manufacturing, trade, service or government) depending on the father's industry. The four groups of respondents were compared for their ratings of manufacturing, trade, service and government industries. The percentage of ratings above or below the industry's composite mean for each group was computed. A chi square test was used to measure the group differences.

TABLE 2

SAMPLING SCHEDULE

	College of Commerce		College of Arts and Sciences		
Form	Men	Women	Men	Women	Sums
A	40	20	40	20	120
B	40	20	40	20	120
C	40	20	40	20	120
	120	60	120	60	360

The students were also compared to ascertain whether the industry they hope to enter influenced their ratings. The procedure was the same as above with father's industry.

Army, Navy, and Air Force veterans as well as nonveterans were compared to discover their agreement on rating the Army, Navy, and Air Force. Other services were not used since the number of veterans in the sample was inadequate. A chi square test was employed to compute the differences.

To evaluate the influence of occupations on the prestige of the industries, representative occupations from the North-Hatt scale were selected. Twenty-two of the industries had occupational correlates on the scale. Since a number of the industries had more than one related occupation, the North-Hatt ranks were averaged to derive a mean rank. The 22 industries and occupations were reranked and a rank-order correlation was computed.

The earlier industrial prestige study conducted by Brayfield et al. was compared with this study. A rank-order correlation was computed between the 22 industries common to both studies. Since the Brayfield study did not report a combined rank order for both men and women, correlations were computed for each sex.

RESULTS AND DISCUSSION

The results are presented in Tables 3, 4, and 5. Table 3 presents the composite rankings based on the mean ratings for each industry. The weighted ratings ranged from a high of 92.2 for Medical Services to a low of 30.1 for Coal Mining. Although the hierarchy shows good dispersion, it should be noted that the score distance between several of the industries was small. For example, Drug Products Manufacturing and Religious Organizations (difference of 0.8); Real Estate Com-

TABLE 3

RANK ORDER OF 52 INDUSTRIES BASED ON MEAN SCORES OF PRESTIGE RATINGS BY 360 COLLEGE STUDENTS

Industry	Rank	Mean Score	Industry	Rank	Mean Score
*Medical services	1	92.2	Marines	27	63.0
Colleges and universities	2	90.1	Weapons mfg.	28	61.8
Research laboratories	3	88.5	Rubber products mfg.	29	61.7
Atomic energy plants	4	87.8	*Local government	30	61.5
*Banking	5	85.0	Department stores	31	60.0
Private schools (secondary)	6	79.4	Ceramic products mfg.	32	59.6
*Commercial airlines	7	79.3	Motels	33	58.0
Television broadcasting	8	77.4	*Furniture mfg.	34	57.7
*Chemical products mfg.	9	77.0	Clothing mfg.	35	57.2
*Aircraft mfg.	10	75.57	Restaurants	36	56.73
*Motion picture industry	11	75.50	Tobacco mfg.	37	56.71
*Federal government	12	75.2	Coast guard	38	56.5
*Drug products mfg.	13	73.8	*Building construction	39	56.2
Religious organizations	14	73.0	*Drug stores	40	55.6
*Publishing companies	15	72.7	Forestry	41	55.3
Insurance companies	16	70.3	Iron and steel mills	42	54.5
Petroleum refining	17	69.4	*Railroads	43	54.1
*Real estate companies	18	66.7	Interstate bus transportation	44	51.7
*Automobile mfg.	19	66.6	Army	45	49.9
Air Force	20	66.4	*Farming	46	47.8
*Telephone companies	21	64.8	Grocery stores	47	47.0
Public schools (secondary)	22	64.7	Meatpacking	48	45.2
*Hotels	23.5	64.4	Labor organizations	49	42.9
Navy	23.5	64.4	*Laundries and dry cleaning	50	39.66
Ship and boat building	25	63.8	Local bus transportation	51	39.60
*Electric power and light companies	26	63.7	*Coal mining	52	30.1

* Denotes industries used by Brayfield et al.
Note.—Mean rating: 63.6. Standard deviation: 13.56.

panies, Automobile Manufacturing and Air Force (0.3); and Weapons Manufacturing, Rubber Products Manufacturing, and Local Government (0.3). In light of this it would be inadvisable to adhere rigidly to the mean ranks as sharp discrimination points on the scale. Further, it can be seen that the distance between the industries is neither equal nor large. Consequently, it might be more meaningful to group the industries on the basis of the rating adjectives and mean score range (see Table 4).

Table 5 summarizes the relationships. Most of the trait comparisons resulted in high correlations, which suggests that such influences as sex of the respondent, prior work experience, academic training, and the presentation order of the industries produce only small differences in the ratings.

The moderate agreement between Brayfield's Kansas respondents and the Ohio respondents in this study implies that geographic differences play an important role in

TABLE 4

INDUSTRIES GROUPED BY RATING ADJECTIVE AND SCORE

Rating Category	Score Range	No. of Industries
Excellent prestige	90 and up	2
Good prestige	80 to 90	3
Average plus prestige	70 to 80	11
Average prestige	60 to 70	15
Fair plus prestige	50 to 60	13
Fair prestige	40 to 50	5
Poor prestige	below 40	3

TABLE 5

SUMMARY OF RELATIONSHIPS

Variables	Rho Correlation Coefficient
Men and women	+.95
Prior work/no prior work	+.97
Brayfield/Campbell (men)*	+.47
Brayfield/Campbell (women)	+.55
Commerce/Arts and Sciences	+.96
Within college (Commerce)	+.96
Within college (Arts and Sciences)	+.97
Occupations and industries	+.84
Agreement among forms	+.94 (Kendall's *W*)

* Significant at .05 level of confidence; all others significant at the .001 level.

determining the relative prestige of industries. The industries that showed the most disagreement for both men and women were Farming, Local Government, General Building Construction, Chemical Manufacturing, Wholesale Drug Companies, and Hotels. The largest rank difference occurred for Farming. Brayfield's respondents ranked Farming, Local Government, and General Building Construction higher.

Respondents grouped by their father's industry did not show rating partiality for his industry, nor did the respondents grouped by the industry they hope to enter.

The three groups of military veterans and the group of nonveterans differed significantly for rating two of the three military services (Army and Air Force). The Army veterans showed the most disagreement and had the greatest percentage of ratings below the composite mean for all of the services. The Air Force and Navy veterans, as well as the nonveterans, rated the three services approximately the same and expressed a slight favoring for the Air Force.

Forty-two percent of the respondents selected "it requires intelligence and ability to enter the industry" as the influence that contributes most to the prestige of an industry. A large proportion also selected "social standing" (21.7%) and 21% chose "it benefits humanity" as the primary determiner of the prestige of an industry. The remainder stated that industrial prestige cannot be attributed to any one influence, but that it is due to a combination of influences such as wealth of the industry, benefit to society, intelligence and ability required to enter, and general social standing. The writer feels that prestige is due to a combination of factors and that it is almost impossible to isolate one factor to account for prestige. This is supported by the composite hierarchy which (Table 3) fails to reveal any high correlation with one factor.

The comparison between the prestige rankings of occupations of an industry selected from the North-Hatt study and the ranks of comparable industries in this study indicated a strong agreement (+.84). The high correlation suggests that the respondents were probably influenced by the typical occupations of an industry. For example, medical services, physician; railroads, conductor; coal mining, coal miner; and banking, banker. Some of the respondents commented that they felt they were influenced by the occupations.

CONCLUSIONS

The hierarchy of industries in terms of prestige ratings is not a sharply defined rank by rank continuum. It can better be described as a hierarchy progressing in clusters. Each cluster is composed of industries that are judged to be relatively similar in prestige status.

Respondents compared on the basis of their personal attributes, namely, academic training, sex, father's industry, the industry they hope to enter, and work experience, were homogeneous in evaluating industries for prestige. The only group comparison that did not result in high agreement was military experience. The Army veterans differed from the other veterans and nonveterans and consistently rated the services below the composite mean.

There are a number of contributing influences that determine the prestige of an industry. The influences primarily include the typical occupations of the industry, the benefit it gives to society, the intelligence and ability required of the employees, and general social standing.

The respondents from two different geo-

graphical areas in the United States showed moderate agreement. It appears that the respondents are influenced by the industrial values of their respective areas in evaluating industries for prestige.

REFERENCES

BRAYFIELD, A. H., KENNEDY, C. E., JR., & KENDALL, W. E. Social status of industries. *J. appl. Psychol.*, 1954, **38**, 213–215.

BUREAU OF THE BUDGET, DIVISION OF STATISTICAL STANDARDS. *Standard Industrial Classification Manual.* Washington: U. S. Government Printing Office, 1957.

COUNTS, G. S. The social status of occupations; a problem in vocational guidance. *School Rev.*, 1925, **33**, 16–27.

NORTH, C. C., & HATT, P. K. Occupations ranked according to prestige. NORC Survey No. 244, Sept. 1, 1947, 3–13.

(Received May 18, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

SOME CORRELATES OF THE EMPATHY OF THE WORKERS WITH THE FOREMAN[1]

URIEL G. FOA

Israel Institute of Applied Social Research

It has been suggested that the performance of a group such as a working team is influenced by the extent to which the group members succeed in judging correctly the perceptions and expectations of the other members (McMurry, 1953). For this reason research findings about the factors which influence empathy are of immediate interest to people dealing with practical problems of group organization.

This paper is concerned with the empathy of the worker toward his foreman when one's perception differs from the other's. It reports on an attempt to predict empathic success or failure of the worker from two kinds of factors: source of motivation and the characteristics of the judged perception. Three motivations are considered: the motivation of the worker, the motivation that the worker ascribes to the foreman, and the motivation of the foreman. A foreman, for example, may be motivated to seek reward from management, thus orienting his attention toward the cues coming from management. For another foreman the rewards coming from the worker are more important, and he will therefore orient his attention toward the worker. The assumption is that one cannot pay attention to all the cues to which one is exposed in an interpersonal situation and that he will orient his attention toward the source which is more meaningful to him in terms of potential rewards (Walters, 1958). It is suggested that the empathic response will be influenced by these differences in orientation. The results of a laboratory experiment described by Lundy (1956) show that the ability to judge the perception of another person improves when the attention of the empathizer is oriented toward this person.

Another possible factor related to success of empathy is the characteristics of the predicted perception. Baker and Block (1957) have found that, in a laboratory experiment, empathic success is more likely when the behavioral description to be predicted possesses socially acceptable traits.

METHOD

The Test of Foreman-Worker Relationship (Foa, 1957) was administered to 361 Israeli factory workers and their 51 foremen. This test consists of two sets of pictures of eight situations each. One set is concerned with the behavior of the foreman, and the other with the behavior of the worker. Both sets of situations were presented to the foreman and the worker. For each situation they were instructed to choose, out of four given types of behavior, the one which (*a*) ought to occur in an ideal situation, and (*b*) the one which occurs most often in practice in their department. Further, each worker was requested to guess the alternatives chosen by his own foreman, both at the ideal and the actual levels.

Scoring. (1) Real Similarity Score: number of times the worker and his foreman gave the same response to the eight situations of a given set of pictures. (2) Empathy Score: number of times the worker succeeded in judging correctly the responses given by the foreman to the pictures of a given set.

Since there are eight pictures in each set, it is evident that the maximum score is eight and the minimum score is zero.

For each worker four empathy scores were obtained, referring to four *types of action,* namely, the ideal behavior of the foreman (IF), the ideal behavior of the worker (IW), the actual behavior of the foreman (AF), and the actual behavior of the worker (AW). Likewise four real similarity scores, one for each type of action, were obtained for each worker-foreman dyad. The present analysis is concerned with workers having a real similarity score lower than 4, for a given type of action. The perception of these workers differs, in most cases, from the perception of their foremen.

[1] This paper is an outgrowth of a project on human factors in production initiated with partial support from the Ford Foundation. The research reported here has been sponsored by the Air Force Office of Scientific Research of the Air Research and Development Command, United States Air Force, through its European Office, under Contract AF 61(052)-121. Reproduction in whole or in part of this paper is permitted for any purpose of the United States Government.

Thanks are expressed to Judith T. Shuval for valuable comments on a preliminary draft of the manuscript.

Within this group of workers, two subgroups are compared: (*a*) Workers who, in most cases, *fail to empathize,* (empathy score below 4). (*b*) Workers who, in most cases, *succeed* in empathizing (empathy score above 4). The empathic outcome thus defined constitutes the predicted variable.

One of the predictors, motivation, was derived as following: Each respondent was asked to indicate, through the choice of appropriate pictures of the test, the source of the motivation of his behavior and of the behavior of the other. One of these sources is the workers' group. When this group was chosen at least twice out of four possible times, the behavior was said to be worker-oriented. Otherwise, it was said to be non-worker-oriented.

The second predictor refers to the characteristics of the perception of the foreman. Real similarity and empathy scores were obtained by comparing this perception, respectively, with the perception of the worker and the perception that the worker ascribes to the foreman. The scores indicate whether the compared perceptions are alike or different.

The design of the test enables the investigator to determine which facets of behavior are different in the two compared responses. Three facets are considered: (*a*) *focus:* behavior can be either instrumental (focused on the job) or affective (focused on the personal relationship); (*b*) *activity:* behavior can be active (doing something) or passive (refraining from doing something); (*c*) *direction:* behavior can be positive, directed toward the focus (fostering the group goal or strengthening the personal relationship), or negative, away from the focus (Foa: 1955, 1958b). Aggressive behavior, for example, is focused on the personal relationship, active and negative. It was therefore possible to observe whether the perception of the foreman, the perception of the worker, and the perception ascribed by the worker to the foreman were alike or different in the value of a particular facet. Further, it was possible to find out whether the success of the worker in judging the perception of the foreman was related to the values of the facets of such perception.

Results

The finding reported in Table 1 concerns the relationship between motivation and the worker's success or failure in empathizing the foreman.

Motivation of the Worker. When the worker is oriented toward his peer group he is somewhat less likely to succeed in empathizing the foreman. This supports the findings of Lundy (1956), but in our case, the percentage difference is small and not significant statistically.

Motivation of the Foreman. When the foreman is oriented toward the workers, the

TABLE 1

Percentage of Workers Succeeding in Empathizing by Type of Action and Motivation

Motivation	Type of Action: Ideal of Foreman	Ideal of Worker	Actual of Foreman	Actual of Worker	All Types
of worker					
worker-oriented	11 (18)	18 (28)	19 (43)	3 (58)	12 (147)
not worker-oriented	13 (24)	21 (19)	20 (40)	8 (38)	15 (121)
Difference	2	3	1	5	3
of foreman					
worker-oriented	20 (10)	33 (15)	22 (37)	14 (35)	21 (97)
not worker-oriented	9 (32)	12 (32)	17 (46)	0 (61)	9 (171)
Difference	11	21	5	14	12*
ascribed to foreman					
worker-oriented	0 (19)	16 (32)	13 (31)	1 (68)	7 (150)
not worker or oriented	22 (23)	27 (15)	23 (52)	14 (28)	21 (118)
Difference	22	11	10	13	14*

* Difference significant at the .01 level.
Note.—The figure in parentheses indicates the total absolute frequency for the given group of workers.

TABLE 2

PERCENTAGE OF INSTANCES OF EMPATHIC SUCCESS OF THE WORKER BY TYPE OF ACTION AND FACETS OF FOREMAN'S PERCEPTION

Facet and Value of the Foreman's Perception		Type of Action			
		Ideal		Actual	
		of Foreman	of Worker	of Foreman	of Worker
Direction:	positive	26 (42)	40 (35)	30 (271)	31 (45)
	negative	4 (79)	4 (72)	0 (66)	0 (277)
	Difference	22*	36*	30*	31*
Focus:	job	15 (33)	20 (69)	29 (141)	14 (109)
	relations	14 (92)	11 (70)	13 (187)	2 (190)
	Difference	1	9	16*	12*
Activity:	active	19 (47)	19 (93)	22 (176)	9 (114)
	passive	10 (100)	11 (80)	17 (154)	2 (166)
	Difference	9	8	5	7*

Note.—Figures in parentheses indicate the total absolute frequencies of dissimilarity cases. These frequencies differ according to different facets. In this Table each instance of success is counted as a unit, while in Table 1 each worker is counted as a unit. Hence the difference in reported frequencies between the two Tables.

worker is more successful in empathizing him. For all the four types of action combined, 21% of the workers succeed when the foreman is worker-oriented, as against 9% when the foreman is not worker-oriented. The difference, 12%, is significant at the .01 level.[2]

Motivation Ascribed to the Foreman. While worker-orientation in the foreman facilitates the empathy of the worker, the orientation that the worker ascribes to the foreman has precisely the opposite effect. When the worker thinks that the foreman is worker-oriented, he is less likely to succeed in empathizing him than when he thinks that the foreman is not oriented toward the worker. For all the four types of action combined, 7% of the workers who perceive the foreman as worker-oriented succeed in empathizing, as against 21% when the foreman is not perceived as worker-oriented. The difference, 14%, is significant at the .01 level. Here, too, as for the orientation of the foreman and of the worker, the differences for each one of the four types of action are in the same direction.

In summary, the empathic judgment of the worker is facilitated when the foreman is oriented toward the workers and/or when the foreman is perceived by the worker as *non*-worker-oriented. Worker-orientation in the worker may also be a hindrance.

Worker's Empathy and the Perception of the Foreman. The results obtained by Baker and Block (1957) suggest that it should be easier for the worker to judge correctly certain perceptive or ideal images of the foreman than to judge certain other images. Three component facets of these images are supplied by the test with two alternative values for each facet: (*a*) focus of behavior (the alternative values of this facet are: toward interpersonal relations or toward the job); (*b*) direction of behavior (toward the focus or away from it); and (*c*) activity of behavior (active or passive). Let us assume that the perception of the foreman is different in a particular facet from the perception of

[2] Empathy scores for different types of action and responses to different situations of the same type of action are positively correlated among themselves. Since these correlations were disregarded in the computation of the standard error of the difference between percentages, the results of the significance tests given here are conservative, i.e., somewhat biased in favor of the acceptance of the null hypothesis.

the worker, the question is then: Is it easier for the worker to empathize correctly when the perception of the foreman is of positive direction than when it is of negative direction? The same question occurs, of course, for the two other facets of behavior, focus and activity. In Table 2 the values of the facets of the perception of the foreman are plotted against the percentages of instances of empathic success of the worker. The table shows that:

(*a*) The percentage of empathic successes of the worker is always higher when the foreman's perception is positive (first row of the table) than when it is negative in direction (second row). All the percentage differences between the positive and negative values of the direction facet are significant at the .01 level.

(*b*) The percentage of successes is higher when the perception of the foreman is focused on the job rather than on relations, but the percentage difference is significant, at the .01 level, only for the actual level of action.

(*c*) The percentage of successes is higher when the perception of the foreman is active rather than passive, but the difference is significant, at the .01 level, only for the actual action of the worker.

In summary, it seems that when the foreman perceives the action as positive, focused on the job, and active, it is easier for the worker to judge correctly the foreman's perception.

Discussion

The results suggest that the motivational orientation of the person whose perception is predicted, the orientation ascribed to him by the predictor, and the characteristics of the predicted perception are all related to success of prediction. While all these factors seem to contribute to the outcome of the empathic process, none appears to have a decisive influence upon it. Perhaps it should be expected that a highly complex process such as interpersonal prediction should be mildly related to a relatively large number of factors rather than being dependent upon very few decisive conditions (Foa, 1958a). It is, however, of some interest that the results obtained here in the field seem to confirm and extend some previous findings obtained in experimental situations.

The present data support to some extent the earlier findings of Lundy (1956) that the orientation of attention of the judge is related to empathy. Furthermore, they show that the attention of the other person, the foreman, also plays a considerable role. More intriguing is the finding that the worker's feeling that the foreman is worker-oriented is likely to reduce the worker's chances for a correct prediction. This may be explained by suggesting that such a feeling may indicate satiation in the worker with regard to the cues coming from the foreman (Gewirtz & Baer, 1958). A satiated worker seems less likely to be perceptive toward the foreman, and his ability to judge the perception of the latter may be consequently reduced.

These findings suggest some practical conclusions. It seems that understanding is increased when the foreman is worker-oriented but succeeds in avoiding overidentification with the worker's group. The ability of the foreman to assume such an attitude may, in turn, be influenced by the attitude of management towards him. If the relationship with management requires the whole attention of the foreman it may be quite difficult for him to orient himself towards the workers. At the same time, if the foreman feels that his status is insecure, he may look to the workers' group for support, thus producing satiation in the worker. A management interested in promoting good communication between worker and foreman will, therefore, avoid monopolizing the orientation of the foreman, while providing him with support. The creation of conditions enabling the foreman to perform adequately as a link between management and workers seems, therefore, to be primarily a responsibility of management.

The finding that the facets of the perceived behavior are related to the success of the empathy confirms the earlier results of Baker and Block (1957), obtained under laboratory conditions. They suggest that empathy is selective as much as perception is selective (Bruner & Postman, 1957). Such selectivity may be a result of previous learning.

SUMMARY

This paper presents some findings regarding the ability of the worker to predict the perception of his own foreman.

The results suggest that the prediction of the worker is more likely to succeed when the foreman's attention is oriented toward the worker. Success is also more likely when the worker believes that the attention of the foreman is not oriented toward the worker. It is also shown that the characteristics of the perception to be predicted significantly influence the outcome of prediction. Thus, certain perceptions are easier to predict than others.

REFERENCES

BAKER, O., & BLOCK, J. Accuracy of interpersonal prediction as a function of judge and object characteristics. *J. abnorm. soc. Psychol.*, 1957, **54**, 37–43.

BRUNER, S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.

FOA, U. G. Foreman-worker interaction; a research design. *Sociometry*, 1955, **18**, 226–244.

FOA, U. G. A test of the foreman-worker relationship. *Personnel Psychol.*, 1957, **9**, 469–486.

FOA, U. G. Empathy or behavioral transparency? *J. abnorm. soc. Psychol.*, 1958, **56**, 62–66. (a)

FOA, U. G. The contiguity principle in the structure of interpersonal relations. *Hum. Relat.*, 1958, **11**, 229–238. (b)

GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditions. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172.

JARRARD, L. E. Empathy: The concept and industrial applications. *Personnel Psychol.*, 1956, **9**, 157–167.

LUNDY, R. M. Assimilative projection and accuracy of prediction in interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, **52**, 33–38.

MCMURRY, R. M. Empathy: Management's greatest need. *Advanc. Mgmt.*, 1953, **18**, 6–11.

WALTERS, R. H. Conditioning of attention as a source of autistic effects in perception. *J. abnorm. soc. Psychol.*, 1958, **57**, 197–201.

(Received January 9, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

THE CONSISTENCY OF RATINGS BY PEERS[1, 2]

DONALD W. FISKE

University of Chicago

AND JOHN A. COX, JR.[3]

Personnel Laboratory, Wright Air Development Center

The way in which a man is perceived by his peers is of considerable practical and theoretical importance. In the armed forces, peer status scores are used as parts of grades, as promotional considerations, and as the basis for a variety of personnel decisions. However, little is known about some of the contextual factors which might affect rating scores. One of the goals of the present study was to gather information about the effects which some variables may have on peer ratings.

More specifically, this study is concerned with the extent to which ratings are affected by the:

(1) kind of task assigned to the group
(2) particular group to which a man is assigned
(3) number of sessions the group has been together
(4) nature of the trait being rated
(5) role of the rater (peer or observer).

PROCEDURE

The basic experimental unit of this research was a group of six *Ss* brought together for a 2-hr. period and assigned a task. Toward the close of the period, each *S* rated his five peers and himself on seven traits. At the same time, these six *Ss* were also rated by six other men who had observed the group throughout the period.

The same group met together four times over an eight-day period. Then each group was dissolved and *Ss* were regrouped. Each *S*'s associates in the second group consisted of five peers with whom he had not worked previously. Each new group also performed four tasks.

At this point, the men who had been *Ss* became observers and the observers became *Ss*. The process was repeated, with the new *Ss* working together for four sessions in each of two groups.

Tasks

Each grouping of six *Ss* worked on four tasks, two physical and two verbal. Each set of four tasks consisted of two verbal alternating with two physical: half the groups had the order P, V, P, V, and half had the order V, P, V, P. Examples of the physical tasks are: assembling a set of dowels and blocks to form a specified structure; rolling a ball up an inclined spiral plane; assembling boards to form a bridge. Examples of the verbal tasks are: making recommendations on national problems; playing the game of "Twenty Questions" as a group; playing "Intercept," a war-game problem.

The tasks were rotated so that each man worked on eight tasks and observed eight different tasks. When a new grouping of *Ss* was made, each *S* had had experience with the same set of four tasks in his previous group.

The experimental sessions were held in classrooms.

Observers

All six *Os* observed and rated all six *Ss*. However, each *O* was the "primary observer" for one *S*. He was instructed to watch him most closely. Every 15 min., he rated this one *S* on his activity level and made notes about what his *S* had done during that quarter-hour and about his impressions of *S*. The principal function of this procedure was to maintain *O*'s attention.

When the six *Ss* were regrouped, the primary *O* followed his *S* to the new group. Thus the primary *O* observed the same man for eight sessions but observed other *Ss* for only four sessions.

Subjects

The basic design called for 144 men. In each of the two replications of the design, six groups of six

[1] The research reported in this paper was supported in whole or in part by Personnel Laboratory, Wright Air Development Center, Air Research and Development Command, United States Air Force, under Contract No. AF 41(657)-157. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Some tables omitted from this paper, the full rating scale, and complete descriptions of the tasks are available in Fiske, Cox, and van der Veen (1959).

[2] We gratefully acknowledge the assistance of Ernest C. Tupes, who contributed to the formulation of the design, and the assistance of Kenneth Howard and William Rechenberg, who helped with developing the tasks, running the groups, and analyzing the data.

[3] Now at Trinity College, San Antonio.

men were observed by 36 observers. As indicated above, the *O*s and *S*s reversed roles after two sets of four sessions each.

A few men failed to complete the entire series and had to be replaced. It was possible, however, to obtain essentially complete data on 138 men.

The *S*s were students at the University of Chicago recruited by advertisements. They were paid for each session. Attendance was maximized by exhortation and by offers of bonuses for those completing the entire schedule.

Their mean age was 20.7 years with a range from 17 to 29. Their mean education was 2.6 years of college with a range from just entering college to three years of graduate work.

Before the group sessions began, *S*s were given a battery of psychological tests in three testing sessions. They were also given a list of all *S*s and asked to indicate those whom they knew. From this information, schedules were meticulously worked out so that the six men assigned to any one of the 48 working groups had no previous acquaintance with each other, and so that the six *O*s had no previous acquaintance with any of the six men they rated. The scheduling could not be perfect: in a few instances, an *S* was known slightly by one of the *O*s who rated him, but never by his primary *O*.

Traits

Of the seven rated variables, four were evaluative and referred to the rater's over-all impression at the time of rating: the rater was asked what he thought the ratee was like, in general, without specific reference to the observed behavior.

The rating of *Liking* (No. 1) was designed to get at the rater's general evaluation of the man. (*S*s did not rate themselves on this trait.) *Leadership Potential* (No. 4) was included because of its importance in military personnel work. *Maturity* (No. 2) and *Breadth of Interests* (No. 3) were selected to approximate two factors which have been found in analyses of peer rating scales.

The other three were descriptive of the *S*'s behavior in each session: 5. *Number of Useful Ideas*, 6. *Considerateness*, and 7. *Assertiveness*. For these, the rater was instructed to rate on the basis of what the man did in that particular session. The descriptive traits were adapted from the three factors which Carter (1954) has found to recur in analyses of ratings on behavior in groups: "aiding attainment by groups," "sociability," and "individual prominence and achievement." The findings of Borgatta, Cottrell, and Mann (1958) lend further support to the selection of these traits.

Analyses

The approximately 100,000 ratings were punched into IBM cards and most of the computational work was carried out on IBM machines.

The basic unit was the score for an *S* on a trait for a session. Three such scores or basic ratings were obtained: (*a*) the mean of the ratings by a man's five *peers* in his working group; (*b*) the rating by the *primary observer;* (*c*) the mean of the ratings by the other five *observers*. (Whenever the unqualified term "observer" is used, it refers to these five raters.) The primary observer's rating was excluded from the *O* mean so that the latter would be comparable in every respect except role with the peer mean.

Most of the analyses were carried out for each of these three ratings on each of the seven traits separately.

Results

Effect of Group

Findings on the effect of group will be reported first because this effect proved to be a major one, which has to be considered when studying other effects.

The basic analysis for the effect of group was the intercorrelation (over 138 men) of the mean rating per trait for each of the eight sessions, arranged in temporal order. Thus the intercorrelations within each of the two groups (Sessions 1–4 and Sessions 5–8) could be compared to the intercorrelations between groups. The medians are reported in Table 1.

It is obvious that the correlations involving the same group are substantially higher than those between different groups, for the peer and *O* ratings. In fact, for Traits 1–4, there is no overlap between the distributions of the correlations. The effect of "group" could be attributed to changes in *S*s' behavior as a function of working with different peers or to changes in perceptions as a function of change in raters. A glance at the correlations for the primary observer, who was the same rater for both groupings, indicates that the "group" effect is primarily a function of the change in raters: when the rater is the same, the several ratings for sessions in the second group correlate as highly with ratings for the first group as the latter do with each other.

Effect of Task

To assess the effect of type of task, each basic rating was identified by the type of task involved, and secondarily by the group (first or second grouping for that man) and the serial order (first or second task of that type within the series for the group). These ratings were then intercorrelated to indicate the relationship between each pair of sessions so identified.

These correlations obviously involve the

TABLE 1

MEDIAN INTERCORRELATIONS BETWEEN SESSIONS INVOLVING THE SAME AND DIFFERENT GROUPS

	Traits						
	1	2	3	4	5	6	7
Peer ratings (N = 138)							
First group	.70	.74	.70	.71	.55	.55	.62
Second group	.74	.76	.71	.75	.56	.69	.63
First vs. second	.23	.35	.24	.33	.29	.27	.40
Observer ratings (N = 138)							
First group	.65	.71	.71	.76	.57	.56	.71
Second group	.65	.68	.68	.65	.48	.58	.58
First vs. second	.10	.26	.33	.37	.33	.11	.51
Primary observer ratings (N = 108)							
First group	.54	.51	.46	.57	.33	.31	.49
Second group	.69	.71	.59	.65	.52	.47	.56
First vs. second	.58	.52	.45	.55	.41	.37	.44

same data as those in Table 1, but combined in a different way. For example, the first physical task encountered by each *S* was the first task chronologically for half the *S*s and the second chronologically for the other half.

Inspection reveals no systematic tendency for the correlations between sessions involving different types of task to be any lower than those between the same type of task. In other words, the consistency of peers' perceptions of *S*'s behavior in different sessions is not affected by the type of task (physical vs. verbal).

Of some interest is the effect of type of task on mean ratings. For verbal tasks, the means of the peer ratings are higher on all traits, while the means for *O* ratings are higher on all traits except Traits 1 and 6 (the means for primary *O* are higher only on Traits 3, 4, 5, and 7). The differences are trivial except for Trait 7 (*Assertiveness*) and Trait 5 (*Number of Useful Ideas*).

The major psychological difference between the physical and the verbal tasks was in the nature of the interpersonal interactions. In verbal tasks, everyone interacted continuously with everyone else: each person was either speaking or listening to a peer.

In contrast, *S*s could not sit for most of the physical tasks. Much of the time they moved around, collecting and arranging materials. It would appear that the greater amount of substantial interaction contributed to the higher, more positive ratings in the verbal tasks, and especially for Traits 5 and 7 which apply to aspects of overt actions visible to the whole group.

Effect of Time

Correlations were computed between the sessions in temporal order. The correlations among the four sessions of one group are generally of about the same magnitude: no major effects are present.

There is a slight tendency for the correlations between the third and fourth sessions in each group to be the highest ones in each block, but this trend occurs chiefly in the peer and observer correlations for the first four traits where the correlation between the third and fourth sessions always exceeds the correlation between the first and second by a small amount. No other general trend emerges. While it might be expected that temporally adjacent sessions might show elevated correlations, such does not seem to be generally true.

In the mean ratings on each trait for each temporal session, no major trends are evident.

Effect of Trait

Are the several effects previously examined different for different types of traits? Table 1 shows a clear difference. With but one excep-

tion, the correlations between sessions in a given group are higher for evaluative traits (1–4) than for descriptive (5–7). Thus it appears that these general evaluations are more consistent across sessions than these specific descriptions of *S*'s behavior. This effect disappears in the between-group evaluations, where the general impressions of different raters agree no more highly than do their descriptions of behavior.

The same general picture appears when the sessions are grouped by task. However, in both analyses, primary *O*s' median correlations for evaluative traits are consistently larger than those for descriptive traits in the cross-group comparisons. A primary *O*'s general impression was affected very little by whatever changes in *S*'s behavior occurred when he was placed in a different group. Hence the change in raters (for peer and *O* ratings) leads to a change in description of an *S*'s behavior and to an even larger change in evaluations.

The a priori distinction between evaluative and descriptive traits was reinforced by the instructions on the rating scale. Nevertheless, the several traits showed two groupings which cut across this dichotomy. In several analyses, the data for *Considerateness* were more similar to those for *Liking*, *Maturity*, and *Breadth of Interests* than to the data for the other two descriptive traits. Ratings on these four traits required indirect inference from the observed behavior.

In contrast, *Assertiveness* and *Number of Useful Ideas* involved judgments or evaluations of observed behavior. Apparently *Leadership Potential* was also rated rather directly from observation, since its data resemble those for these two descriptive traits.

The means for these latter three traits tend to be lower and the standard deviations higher than the corresponding values for the four more inferential traits. Again, peers and *O*s agreed less well on the four inferential traits. (Further evidence on trait differences is given in Fiske [1960], and van der Veen and Fiske [1960].)

Discrepancy between Rater and Group Mean

The extent of agreement between the mean ratings and those made by individual raters was explored. For each peer, the absolute discrepancy between his rating and the mean peer rating was calculated for each rating he made on each trait on each other peer in each situation. These were averaged over the seven traits to obtain his discrepancy for each *S* in each session. For each session, the mean and the variance of his discrepancies for the other five *S*s were computed. Comparable analyses were made for *O* ratings.

Each rating was, on the average, within one-and-a-half steps (on the 10-step scale) of the mean of four or five other raters. This mean discrepancy suggests that somewhat different frames of reference existed in the different raters, and that for a reliable rating by peers, a number of raters should be averaged.

The analyses of peer discrepancies were based on 48 men who participated in eight sessions in which no peer was absent. It was deemed advisable to use a smaller sample for whom complete data were available than to use more cases with many estimated values.

An analysis of variance of the mean peer discrepancy scores per session yielded a significant interaction ($p < .01$) between rater and group (first vs. second). Thus the size of a peer's mean discrepancy depended upon the particular group with whom he was working. When the mean square for raters was tested against the mean square for interaction (Rater × Group), the F was barely significant at the .05 level. The presence of individual differences in discrepancy scores was no surprise because individual differences in raw ratings had been noted in inspecting the original data. Since individual differences between raters existed, it seems reasonable to interpret the significant interaction in terms of rater variation, rather than ratee variation: a rater's discrepancy from a group mean is a function of the raters with whom he is compared. The people being rated probably make much less difference.

No significant effect for group (first vs. second) was found.

In a separate analysis of variance, the mean discrepancies were ranked in chronological order within each group. No effect of serial order was found.

There was no mean difference between peers who had been in the role of observers previ-

ously and peers who had not: mere experience in making ratings did not affect the average discrepancies.

The variance over each man's five peers was also analyzed. Since the variances were badly skewed, a nonparametric technique using chi square was employed (Mood, 1950, Ch. 16). The interaction (Rater × Group) was again significant but the rater and group effects were not significant by themselves.

Effect of Role

In the analyses presented above, few differences between peer ratings and *O* ratings have been found. One effect associated with the peers' personal interaction with each other is evident in the mean ratings. On every trait, the mean peer ratings are above the mean for observers or primary observers. (The means over all traits are 5.54, 5.11, and 5.03, respectively.)

The primary observer ratings seem to resemble the observer ratings. No effects seem to be associated with the primary *O* role which cannot be explained by the lower reliability of one rating (as against the mean of five) and by the fact that, in the second grouping, this man saw the same man he had seen before and about whom he had had time to develop a rather firm impression.

The agreement between peer and *O* ratings was investigated. For each of the 192 individual sessions, the peer means and the *O* means were ranked over subjects on each trait. These ranks were correlated, using rho.

The 1,334 values were summarized by trait, and subdivided in various ways. No significant effect was found for group (first or second) or for cycle (having had previous experience in the opposite role or not). Thus again, no effect of experience or time appeared.

For each type of task, a study of trait differences was made by ranking the median correlations of the seven traits within each specific task. Significant trait differences ($p < .001$) were found in each case, using an analysis of variance of ranks.

The median correlations for each type of task and for all sessions are given in Table 2. The medians show essentially the same order (rho = .78). As discussed earlier, the trait differences are presumably associated with the relative specificity of behavioral referents in the trait descriptions. Of most interest are the high agreement on *Assertiveness* and the comparatively poor agreement on *Considerateness*.

TABLE 2

MEDIAN CORRELATIONS BETWEEN PEER AND OBSERVER RATINGS FOR SEPARATE SESSIONS[a]

Trait	Physical	Verbal	All
Liking	.40	.46	.44
Maturity	.54	.55	.54
Breadth of interests	.57	.53	.54
Leadership potential	.67	.72	.70
Number of useful ideas	.75	.70	.73
Considerateness	.40	.59	.49
Assertiveness	.83	.86	.84

[a] The medians are based on rank correlations for each of 192 sessions, half of each kind. For each session, $N = 6$.

The trait differences are more important than the general level of the medians. The agreement between raters is largely a function of the trait being rated.

The degree of agreement between peers and *O*s can also be studied by comparing different sessions. It was found that ratings by peers and *O*s from different sessions agreed better for *Assertiveness* than for *Maturity* and *Considerateness*.

Differentiation of the Traits

The analyses reported above have determined the influence of various factors upon the peer ratings. This section will consider the effect of general impressions as contrasted to specific trait designations.

For the mean ratings in the second grouping, intercorrelations were run for peers and for *O*s separately. For each role, the four evaluative traits had intercorrelations between .69 and .85. The other values were generally lower. The median for all correlations was .70; from this, it may be estimated that the median intercorrelation for ratings made in a single session would be .37.

The generally high level of these intercorrelations among traits accounts for much of the essential homogeneity of the results for the several traits as reported in earlier sections (cf. Table 1). On the other hand, the differ-

ences between the peer–observer correlations for the several traits (cf. Table 2) become even more striking since they emerge in spite of this high overlap between the traits.

Another question is the extent to which the peers and *O*s could jointly identify and differentiate several traits. To study this question, the mean ratings of peers and of *O*s for each group were intercorrelated on four traits, yielding a multitrait–multimethod matrix (Campbell & Fiske, 1959) in which the sets of peers and *O*s in the two separate groups represent four independent "methods."

There is reasonable agreement between "methods" when the peers and *O*s are observing the same behavior, or when ratings from the same role (peer or *O*) but different groups are correlated. For these comparisons, the differentiation of traits is fair. However, when peers in one group are correlated with *O*s in another group, we find lower agreement and much poorer differentiation. In all comparisons, the differentiation between *Leadership Potential* and *Assertiveness* was particularly bad.

Thus dissimilarity between raters in identity and role and differences between the behavior-samples observed are factors contributing to lower agreement between ratings on a given trait and to increasingly poorer discrimination between traits.

Discussion

From the results, one general finding emerges clearly: the rating given an *S* is influenced considerably by the particular men making it. Furthermore, this effect is probably a function of the different frames of reference used by different raters. Thus the individual differences in mean rater discrepancies seem to account for the effects associated with group.

In all probability, this individual rater effect is a function of his average rating. While *S*s as a sample produced ratings with a mean only slightly above the center of the scale, some *S*s concentrated their ratings in the top steps, while others spread them out over the entire 10 steps.

While *S*s did work seriously on the tasks, the typical level of involvement was not high. None of the tasks involved *S*s' major areas of specialization, so that there was little challenge to prove their competence. Most *S*s did not exert maximum effort, in spite of the prize money offered for the groups performing best.

The general setting which produced this moderate level of involvement in the tasks produced a similar—or lower—level of involvement in the ratings: the *S*s knew that their ratings would have no personal effect on themselves or on the peers they rated.

Thus there were several factors in this experiment which worked together: the consistency of the experimental conditions from session to session; the moderate to low involvement of *S*s; the absence of feedback (the raters did not know how their ratings compared with those made by others). The total effect of these factors probably contributed to the consistency of the ratings over sessions, over types of task, and over roles.

Summary

Men were brought together in groups of six to work on assigned verbal or physical tasks. Each group spent four sessions together. The men were then reassigned to separate groups with which they worked for four more sessions. At the end of each session, each man rated his five peers and himself on seven traits. The four evaluative traits were *Liking, Maturity, Breadth of Interests,* and *Leadership Potential.* Ratings on the descriptive traits—*Number of Useful Ideas, Considerateness,* and *Assertiveness*—were based on what the man did in that session. The men were also rated by six *O*s. New *O*s rated the men in the second group, except for one primary *O* who followed the same man through all eight sessions.

Intercorrelations among the mean peer ratings and among the mean observer ratings per session yielded the following findings: 1. A change in the group with which a rater worked was associated with a change in the ratings he received. (Evidence is presented to support the interpretation that this change in rating was primarily a function of the fact that different raters rated him in the two groups.) 2. The type of task assigned to the group had no important effect on the ratings. 3. The number of sessions the group had been

together had no important effect on the ratings. 4. No general effects associated with rater role as peer or *O* were found. Peer ratings were, however, systematically more positive than those by *O*s.

It was also found that agreement between peers and *O*s on ratings in a single session was highest for *Assertiveness*, high for *Number of Useful Ideas* and *Leadership Potential*, lower for *Breadth of Interests*, *Maturity*, and *Considerateness*, and lowest for *Liking*.

Discrepancies between the ratings made by a man and the mean of ratings made by others in the same role showed reliable individual differences.

It was concluded that individual differences in rater frames of reference contribute to the differences between the ratings a man receives in different groups. In comparison to this effect, such factors as type of task, rater's role, and number of previous sessions have little effect on ratings under the conditions of this experiment, in which there was only moderate involvement on the part of *S*s and no opportunity to compare one's ratings with those made by others. It was also concluded that the degree of agreement between raters is a function of rating instructions and trait definition.

REFERENCES

BORGATTA, E. F., COTTRELL, L. S., & MANN, J. H. The spectrum of individual interaction characteristics: An inter-dimensional analysis. *Psychol. Rep.*, 1958, **4**, 279–319. (Monogr. Suppl. No. 4.)

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

CARTER, L. Evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

FISKE, D. W. Variability among peer ratings in different situations. *Educ. psychol. Measmt.*, 1960, **20**, in press.

FISKE, D. W., COX, J. A., JR., & VAN DER VEEN, F. Consistency and variability in ratings. Lackland Air Force Base, TEX.: Wright Air Development Center, 1959. *Res. Rep.* WADC-TR-59-37.

MOOD, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

VAN DER VEEN, F., & FISKE, D. W. Variability among self-ratings in different situations. *Educ. psychol. Measmt.*, 1960, **20**, in press.

(Received February 18, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

DRIVER BEHAVIOR AND SCORES ON THE MMPI[1]

PAUL L. BROWN

Minnesota State Highway Department

AND RALPH F. BERDIE

Student Counseling Bureau, University of Minnesota

If measurable personality variables are related to behavior of drivers, then persons responsible for training and licensing drivers should consider how this relationship can be used in order to reduce the number of highway accidents. Accidents result from a complex involving, in addition to the driver, the machine, the highway, and driving conditions such as the weather and the time of day. Any systematic programs that will effectively reduce the number of accidents must pay attention to all of these variables, but to date too little attention has been given to the personality of the driver.

In a study of traffic accident repeaters in Michigan and Connecticut (ENO, 1948), the Cornell Word Form, a type of personality test, distinguished to some extent between drivers with and without traffic accidents. Relationships were small. Harris (1950) found no significant differences between accident-prone and non-accident-prone employees using a multiple-choice Rorschach, the Rosenzweig Picture-Frustration Test, and 71 Bernreuter-type items. A military aviation study (O'Gorman & Kunkle, 1947) found some statistically significant and suggestive differences on the MMPI among three groups of pilots with different accident records. Of a no-accident group 12% had *Pd* standard scores of 60 or above and 2% *Ma* standard scores of 75 or above. Of a group with accidents due to violating rules, 20% had *Pd* scores of 60 or above and none had *Ma* scores of 75 or above, but of a group having accidents due to taxiing with carelessness, 16% had *Ma* scores of 75 or above.

The results that are available from studies of relationships between personality and accidents are far from conclusive but are intriguing.

The purpose of this study was to determine, after a period of four to six years, if there is a relationship between the accident and violation records of drivers and their scores on the Minnesota Multiphasic Personality Inventory.

METHOD

The subjects of this study consisted of 993 residents of Minnesota who were registered as freshmen in the General College of the University of Minnesota in 1951, 1952, or 1953. This college offers a two-year program in general education, and the students in the college tend to be comparable in terms of ability and high school performance to the total group of high school graduates in the state. Most General College students come from the Minneapolis-St. Paul area, and most fall within the age range of 18 through 22. The mean age of the group was about 19 at the time the MMPI was administered.

During the three years, approximately 2,000 students were registered as freshmen in this college. Complete information regarding both personality inventory scores and driving records was available for slightly over 1,000, and as relatively few persons for whom such information was available were women, the study was restricted to men. The difficulty of obtaining driving records for women was due to the change of women's names resulting from marriage.

When a student enters the General College, he is expected to take a group of tests, including the MMPI. The proportion of entering students who actually take this test varies from year to year from 70 to 90%. The names of entering students were obtained, the students identified for whom test scores were available, and then these names were checked against the records of the Minnesota State Highway Department. From the driver's license application form in this department, the age, address, and date of issuing the driver's license were obtained. From the driver's record packet in this office was obtained information regarding violations, including date of violation, type of violation, place of violation, and department action taken against the driver, along with information about accidents which included date of accident, number of accidents, and the accident records file number.

[1] This research was done with the cooperation of the Minnesota State Highway Department. The opinions and assertions made, however, are those of the writers and should not be construed as official or representing the views of the Highway Department.

After the frequency distributions of the number of accidents and number of violations were obtained, three groups of 100 individuals each were selected from the 993 males for more intensive study. The first group consisted of the 100 persons with the greatest number of violations and accidents. In this group were persons who had five or more violations and/or three or more accidents. The low group consisted of 100 persons selected from the total group having no violations and no accidents on their records. A middle group of 100 persons was selected who fell at about the mean of each distribution, and these were people who had one and two violations combined with one accident. Frequency distributions were then tabulated, and the mean raw scores on the MMPI were compared for these three groups. The proportion of persons in the low group and the high group answering each question on the *Pd* (Psychopathic Deviate) scale and the *Ma* (Hypomanic) scale was determined, and the groups were compared in order to select items which differentiated between the groups.

A questionnaire was prepared and sent to each person in each of these three groups. Included on this questionnaire were questions asking for the number of miles driven per year, the number of years the person had been driving, the number of accidents the person had had, and the number of traffic tickets other than for parking. Two follow-ups to the original questionnaire resulted in a return of approximately 80%.

TABLE 1

FREQUENCY DISTRIBUTIONS OF NUMBER OF VIOLATIONS AND NUMBER OF ACCIDENTS

Violations		Accidents	
Number of Violations	Number of Men	Number of Accidents	Number of Men
0	321	0	421
1	229	1	229
2	155	2	142
3	107	3	78
4	72	4	27
5	44	5	18
6	18	6	2
7	22	7	4
8	11	8	2
9	5	9	1
10	3		
11	2		
12	1		
13	1		
14	1		
15	1		
Total	993	Total	993

Note.—Mean violations = 1.88. Mean accidents = 1.09.

TABLE 2

MEAN RAW SCORE FOR EACH SCALE OF THE MMPI FOR HIGH, MEDIUM, AND LOW GROUPS

Scale	Mean		
	Low Group	Middle Group	High Group
?	3.35	1.93	.42
L	3.77	3.41	3.27
F	4.58	4.91	5.81
K	15.03	14.82	14.75
Hs	12.40	13.08	12.56
D	17.31	18.54	17.71
Hy	18.70	19.73	19.24
Pd	22.52	23.29	23.99
Mf	22.65	22.58	21.70
Pa	8.50	8.98	9.11
Pt	28.97	29.53	27.28
Sc	28.71	28.60	27.99
Ma	20.47	20.57	21.44
Si	23.64	24.01	22.81

RESULTS

Table 1 presents the number of violations and the number of accidents for the 993 drivers as shown in the Highway Department records. Approximately one-third of the drivers had no violations, more than 10% had five or more violations. More than one-half of the drivers had at least one accident, and well over 10% had at least three accidents. The mean number of violations was 1.9, and the mean number of accidents was 1.1.

It must be noted here that many of these drivers had been driving for at least two or three years before they took the MMPI, and the number of violations and accidents includes all of those reported in Minnesota, some before and some after the time of testing.

Table 2 shows the mean scores on the scales for the three separate groups. In general, the means are not significantly different with the exception of the *Pd* and *Ma* scales, where the poor drivers tend to obtain elevated scores.

The correlations between the number of accidents and the number of violations and scores on the MMPI are shown in Table 3. With a group of 993 persons, a correlation of .08 is significant at the .01 level. Here it is seen that the significant correlations with *Pd*

TABLE 3

CORRELATION COEFFICIENTS FOR NUMBER OF VIOLATIONS AND NUMBER OF ACCIDENTS FOR EACH SCALE OF THE MMPI

Scale	Correlation Coefficient of Scale and Violations	Correlation Coefficient of Scale and Accidents
L	−.051	.007
F	.088	.079
K	.000	−.038
Hs	−.005	.000
D	−.029	−.007
Hy	.038	.002
Pd	.101	.083
Mf	−.040	−.032
Pa	.040	.046
Pt	−.004	−.027
Sc	.001	.000
Ma	.075	.097
Si	−.078	−.006

and *Ma* scales are so small that they have little practical value.

Examination of the profiles suggests that the obtained differences in the group analyses are small because the group of poor drivers contains a number of different kinds of profile groups or personality types. One might speculate, for instance, that one kind of poor driver, perhaps characterized by a high *Pd* score, is the driver who doesn't conform to much of anything, who doesn't have respect for the rights of other individuals or groups, who is quite completely self-centered. Another kind of driver well might be one who is extremely hostile toward other persons, perhaps as shown by a relatively high *Pa* score. His driving behavior may be motivated by a desire to get even with the next driver or to show the next fellow. A third type of driver might be one with an elevated *Ma* score, the driver who is in a great hurry and whose main interest is in reaching his destination as soon as possible, or sooner. When scores for these different types of drivers are combined, the elevated scores of one group tend to be compensated for in group analysis by the depressed scores of other groups, and in general the type of group analysis done here reveals little.

The self-reports of the number of years of driving and the number of miles driven per year were compared for the three groups. A chi-square test showed that the poor driving group reported they had been driving a slightly significant greater number of years than the good driver group, but the difference in terms of number of miles per year was not significant. The number of accidents and violations in the self-reports were compared to those in the Highway Department records, and a strong and consistent tendency was found for drivers to report fewer accidents and violations. The mean number of accidents recorded for the poor drivers was 3.49, the mean number they actually reported themselves was 2.77, an understatement of 21%. The mean number of violations recorded was 5.88, and the mean of the personal reports was 3.98, an understatement of 33%.

CONCLUSION AND DISCUSSION

The results indicate that there is a small but observable relationship between scores on the MMPI and the number of accidents and violations of drivers. The observed differences suggest that a list of test items might provide some assistance in identifying drivers prone to have accidents because of their own personality characteristics.

If persons can be identified who have different types of personalities related to accidents and violations, perhaps the actual occurrence of the accident or violation can be reduced simply by calling to the attention of the individual the fact that he has this pattern of characteristics. Perhaps more intensive instruction would be necessary and it may be that certain persons can be identified, in driver training courses or through the means of driver's license examinations, who require more extensive counseling or perhaps even psychotherapy in order that they understand their own motivations sufficiently to control them when they are at the wheel.

Another possibility is suggested here by the difference in the self-reports and the actual recorded number of accidents and violations. Drivers who tend to under-report their violations may not be lying, they actually may be reporting themselves as they perceive themselves. Most persons perceive themselves as good, competent drivers, particularly young

men in this age range. Driving well is one of the things almost all young men feel themselves capable of doing. Any experience that tends to threaten this self-perception may arouse many defenses, and the most obvious kind of evidence that a person may not be as good a driver as he believes, that is, actual arrest or traffic tickets, may be quickly shut out from awareness or repressed so that the person can maintain the favorable perception he has of his own driving skill. The first problem with persons who have poor driving records may be to get them to realize they are poor drivers.

Adler (1941) proposed on the basis of a clinical study of 130 industrial accident cases five personality types prone to have repeated accidents. These were:

Persons with revengeful attitudes
Persons considering themselves as unlucky
Persons longing to be pampered
Persons who are overly ambitious
Persons who are overly fearful

The concept of a personality type which is accident prone is difficult to defend, but the concept of habitual patterns of attitudes and behaviors which contribute differentially to the occurrence of accidents finds some support in this study.

An attitude of unconcern about others, with a failure to benefit from experience, accompanied by a disdain for rules and regulations, the *Pd* syndromes on the MMPI, well might be related to accidents.

Another set of attitudes more directly involves feelings of hostility. Persons who habitually respond to others with suspicion, who view every social situation as competitive, who strive for dominance in every instance, may as a result of these attitudes have accidents. This is the *Pa* syndrome.

Persons with high *Ma* scores may value effort and movement for their own sake, may pay little attention to social consequences of their acts, and may have accidents.

Not all persons with these attitudes will have accidents, for the results of some of these attitudes can be suppressed by other attitudes. The *Pa* syndrome may be suppressed by a tendency to withdraw from others, or by a fear of injury. None of these speculated relations are simple; rather, they are only suggestive.

Summary

A group of 993 men was administered the MMPI when they were freshmen in college. From four to six years after this, their driving records were investigated, and a very small but statistically significant relationship was observed between scores on the *Pd* and *Ma* scales of the MMPI and the number of accidents and the number of violations of these individuals. The small size of these relationships might be due to the existence of more than one kind of personality pattern related to poor driving. Knowledge of the kind of personality organization and motivation of a driver may be useful for purposes of both licensing and training drivers.

REFERENCES

Adler, A. The psychology of repeated accidents in industry. *Amer. J. Psychiat.*, 1941, **98**, 99–101.

ENO Foundation for Highway Traffic Control. *Personal characteristics of traffic accident repeaters.* Saugatuck, Connecticut: Author, 1948.

Harris, F. J. Can personality tests identify accident-prone employees? *Personnel Psychol.*, 1950, **3**, 455–459.

O'Gorman, W. D., & Kunkle, E. C. Study of the relation between Minnesota Multiphasic Personality Inventory scores and "pilot error" in aircraft accidents. *J. aviat. Med.*, 1947, **18**, 31–38.

(Received February 27, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

COMPANY POLICIES AND SUPERVISORS' ATTITUDES TOWARD SUPERVISION[1]

ERWIN S. STANTON

Bernard M. Baruch School of Business and Public Administration, City College of New York

This investigation was designed to study the relation between company policies and supervisors' attitudes toward supervision in two companies, one of which was characterized as having democratic management, the other, an authoritarian management.

Research Design

Selection of the two companies. Two industrial organizations were selected for detailed investigation because from a preliminary study of anecdotal data, firsthand observations, and a tentative analysis of personnel policies, it appeared that these two companies differed substantially in their approach to management. Both companies are leaders in their respective industries and maintain their executive and administrative offices in a large metropolitan city along the eastern seaboard. The democratic company is a medium sized organization engaged in the metal fabricating industry and operates a number of manufacturing plants in various parts of the country. Its executive office, comprising approximately 70 employees, is primarily concerned with administrative and managerial functions. The authoritarian company is one of the nation's largest operating and holding companies with varied and diversified interests throughout the United States as well as in several foreign countries. Approximately 2,000 persons are employed in the company's corporate office and are engaged in carrying out duties similar to those performed in the democratic company.

Personnel policies in the two companies. In each of the two companies, personnel policies currently in practice were subjected to a detailed content analysis. This was done by studying written statements, administrative decisions, manuals, annual reports, and letters to employees and stockholders as well as stated goals and objectives. In addition to investigating personnel policies appearing in company literature, management officials in both organizations were interviewed using an unstructured interview. All data gathered in this fashion were classified into numerous categories. This was then followed by an investigation of what appeared to be significant differences as well as similarities between the two companies. Although there were fundamental differences in personnel policies between the two companies, there were also numerous basic similarities. Of primary significance was the finding that in *both* companies consideration for the employee was accorded major importance. In both the democratic and the authoritarian companies management officials were most emphatic in stressing the importance of treating their employees with maximum consideration. In addition, both organizations provided favorable working conditions, attractive salaries and advantageous benefits for their employees.

Differences between the two companies. Besides similarities found in both companies, the content analyses of policies of the two organizations disclosed substantial differences which are grouped under six factors:

[1] This paper is based upon a portion of a Ph.D. dissertation submitted to Teachers College, Columbia University. The author wishes to acknowledge the guidance of his advisers, Albert S. Thompson, Robert L. Thorndike, and Kenneth F. Herrold.

Differences between the Two Companies

Factor	Democratic Concept	Authoritarian Concept
(1) Philosophy of Management	An industrial organization must show a profitable operation, but it also has a broad obligation to its employees and to the public.	The function of a commercial enterprise is basically to show a profitable operation. All other considerations are secondary.
(2) Concept of Supervision	A supervisor should give leadership and guidance to his employees and build an efficient and satisfied cooperative work team. This is accomplished by working with and through his subordinates.	A supervisor should develop his employees to the extent that they understand exactly what is expected of them, so that they will be able to perform their work efficiently. Emphasis is placed on the supervisor's own personal qualifications as evidence of leadership ability.

Differences between the Two Companies—*Continued*

Factor	Democratic Concept	Authoritarian Concept
(3) Joint Participation	Employees participate with management in defining the goals and policies of the company as well as take part in the planning of the work of their own departments.	All policy formulation is decided upon by management without employee participation. Work assignments are planned by department heads and then distributed to each employee.
(4) Releasing of Information	Employees are entitled to a great deal of information about the company and its activities. In order to foster the spirit of cooperation, the better informed the employees are, the better they can contribute to the progress of the company.	Information should be restricted to a select group of management officials. Employees need to know only enough information as is required for the successful performance of their duties.
(5) Recognition of the Individuality of the Employee	The employee's individuality is recognized and respected. Persons are not alike in all ways and the law of individual differences is applied in dealings with employees.	Persons are regarded as essentially alike and consequently must be treated in a uniform manner according to established rules and regulations determined by management.
(6) Rigidity-Flexibility	A general flexibility is prevalent. Employee rules, regulations, and policies are more flexibly interpreted, taking individual circumstances into account.	A general lack of flexibility is prevalent. Employee rules, regulations, and policies are rigidly stated, and deviations occur rarely and only under very unusual circumstances.

Raters' evaluation of personnel policies in the two organizations. Because of the unavailability of an established scale of democratic-authoritarian management, it was necessary for the investigator to construct has own scale based on a content analysis of personnel policies in each of the two companies. This introduced a regrettable, but unavoidable, element of subjectivity. To some extent, however, it was felt that this subjectivity could be significantly reduced if the data obtained from the content analyses of the two companies' policies could be evaluated by a group of impartial raters. Accordingly, the data that appeared to differentiate the two companies were summarized into 60 matching statements, half describing the democratic company and the other half referring to the authoritarian company. The statements were then categorized into the six factors previously defined and presented in an analytical and nonidentifying manner to five raters, each of whom was actively engaged in the field of personnel psychology. Without knowing to which company any given item referred, the raters were asked to evaluate the data presented using a 7-point rating scale along a democratic-authoritarian continuum. Average ratings for the 60 statements were then computed and tested for significance. The difference in mean scores between the two companies was found to be significant far beyond the .01 level. Apparently the raters agreed with the investigator that, based on the statements presented to them as a result of the content analyses, the two companies did differ.

Measuring instruments. The two instruments used in this investigation were the *Supervisory Practices Inventory*, Form D (Nelson, 1955) and the *Leadership Opinion Questionnaire* (Fleishman, 1957). The first instrument is an untimed, 50-item, multiple-choice inventory of internalized attitudes toward supervision (Nelson, 1949). This inventory yields a "human relations index" which appears to be somewhat comparable to the Consideration Scale of the *Leadership Opinion Questionnaire*, discussed below. This dimension is referred to in this study as "consideration."

The Leadership Opinion Questionnaire is an untimed, 40-item, multiple-choice questionnaire describing how a person thinks he should operate in a leadership situation. The questionnaire has been factor-analyzed to yield two independent leadership dimensions: "Consideration" and "Initiating Structure." These dimensions have been described as follows:

> Consideration—Items with high positive loadings on this factor were associated with behavior indicative of friendship, mutual trust, respect, and a certain warmth between the leader and his group. High negative loadings appeared on items which suggest that the leader is arbitrary and impersonal in his relations with group members.
>
> Initiating Structure—Items with high positive loadings on this factor imply that the leader organizes and defines the relationship between himself and the members of his group. He tends to

define the role which he expects each member to assume, and endeavors to establish well-defined patterns of organization, channels of communication, and ways of getting the job done (Fleishman, Harris, & Burtt, 1955).

In addition to these two instruments, each *S* completed an autobiographical questionnaire which solicited such data as age, education, years of employment with the company, years of supervisory experience with other companies and the number of employees supervised.

Subjects. A total of 86 *S*s participated in this investigation, 28 being from the democratic company and 58 from the authoritarian organization. All of *S*s were office supervisors of clerical and administrative personnel and were responsible primarily for women clerical employees. Personal characteristics of *S*s were analyzed, and the only significant difference found between supervisors in the two companies was that those from the authoritarian company were responsible for a much larger group of employees than were supervisors from the democratic company.[2]

Results

Company policies and supervisors' attitudes on consideration toward employees. As Table 1 indicates, on neither of the two tests measuring a supervisor's attitude on showing consideration toward his employees was there a significant difference between the democratic and authoritarian companies. Even though the two organizations showed significant differences along the democratic-authoritarian management continuum, their mean scores on the Consideration Scale of the *Leadership Opinion Questionnaire* and on the human relations index of the *Supervisory Practices Inventory* did not differ significantly.[3] There also appeared to be no significant difference between the standard deviations of the two organizations. The data suggest that the attitude of supervisors in companies characterized by authoritarian policies need not necessarily be less considerate of the feelings and desires of their employees than the attitude of supervisors in democratic companies.

It was further interesting to note that in the authoritarian company a positive relation was found between length of service and supervisors' attitudes concerning showing consideration toward their employees. This was indicated by correlations of .41 (significant at the .01 level) and .26 (significant at the .05 level) on the Consideration Scale of the *Leadership Opinion Questionnaire* and the *Supervisory Practices Inventory*, respectively. One possible explanation for this finding is that supervisors in the authoritarian company adopt with the passing of time a more benevolent and paternal attitude toward their subordinates. On the other hand, it is also possible that differences in selection standards may account for these results. Perhaps

[2] Further analysis indicated that in neither company was there a significant relation between the number of employees supervised and supervisors' attitudes toward supervision. Differences in the two companies could apparently not be explained as a function of the size of group supervised.

[3] An analysis of mean scores of supervisors in both companies on the Consideration Scale of the *Leadership Opinion Questionnaire* indicated that the scores were not significantly different from normative data reported by Fleishman, 1957. No normative data is as yet available for Form D of the *Supervisory Practices Inventory*.

TABLE 1

Supervisors' Attitudes on Consideration Toward Employee

Instrument	Democratic Company			Authoritarian Company			Critical Ratio
	Mean	*SD*	Range	Mean	*SD*	Range	
Leadership Opinion Questionnaire Consideration Scale	54.5	6.8	42–65	52.6	8.1	36–70	1.13
Supervisory Practices Inventory	29.4	4.2	23–36	27.8	5.6	16–44	1.47

TABLE 2

SUPERVISORS' ATTITUDES ON IMPOSING OPERATING STRUCTURE

Instrument	Democratic Company			Authoritarian Company			Critical Ratio
	Mean	*SD*	Range	Mean	*SD*	Range	
Leadership Opinion Questionnaire Initiating Structure Scale	50.1	5.5	42–61	56.4	7.0	38–69	4.50*

* Significant at the .01 level.

selection standards used in employing supervisors many years ago were different from those more recently in practice. It is further possible that still other factors yet to be investigated may have been operating and that these could explain the relation between longevity and supervisors' attitudes on showing consideration toward employees. In the democratic company no significant correlation was found in this particular area. It would appear that longitudinal studies need to be carried out before definitive conclusions can be drawn.

Company policies and supervisors' attitudes on imposing operating structure. Table 2 illustrates that there was a significant difference between supervisors in the two companies on the Initiating Structure Scale of the *Leadership Opinion Questionnaire*. Apparently, supervisors in the authoritarian company felt that they themselves should clearly define and facilitate group interaction toward goal fulfillment and provide for comparatively little employee participation in decision making. These supervisors believed that it was primarily their responsibility to initiate leadership functions. Supervisors in the democratic company, on the other hand, believed in imposing far less operating structure in their dealings with employees and favored allowing them greater participation. There was no difference between the standard deviations on this scale for the two companies investigated.[4]

[4] A comparison of mean scores obtained by supervisors in the two companies on the Initiating Structure Scale of the *Leadership Opinion Questionnaire* with normative data published would suggest that both organizations represented opposite extremes on this leadership dimension. The mean score on this scale for the democratic organization was significantly lower and the mean score for the authoritarian company was significantly higher than normative mean scores reported by Fleishman, 1957.

DISCUSSION AND CONCLUSIONS

Research investigators differ greatly in their conclusions as to the type of management that should prevail in an industrial organization. Certain investigators have for some time reported on the advantages of the democratic approach to leadership (Bavelas & Lewin, 1942; Bradford & Lippit, 1945; Gordon, 1955; Viteles, 1953). On the other hand, writers critical of democratic management have recently expressed a preference for a return to a more authoritarian type of leadership (McMurry, 1958; Tannenbaum & Schmidt, 1957). According to supporters of democratic management, where employees are permitted to participate in decision making, morale and productivity correspondingly increase and many of the psychological needs of the individual tend to be satisfied. Critics of democratic management, however, point out that the majority of employees are neither able nor interested in engaging in participative management and suggest that democratic leadership may even impede industrial effectiveness. These two opposing points of view have presented somewhat of a dilemma not only to research investigators, but also to management officials who pose the question: "Under what type of leadership—democratic, authoritarian or perhaps a compromise be-

tween the two—can the organization operate most effectively and meet its objectives?"

Implications of this investigation. This investigation was concerned with studying the relation between policies of a democratic and an authoritarian company and supervisors' attitudes toward supervision. It was found that in *both* companies there was no significant difference in supervisors' attitudes as to the extent of consideration that should be shown for the feelings and desires of employees. There was, however, a very significant difference in the attitude of supervisors as to the type of operating structure that they felt should be established in their dealings with employees. Supervisors in the democratic company believed that they should offer less operating structure and favored permitting employees a greater amount of control over their jobs. In the authoritarian company, on the other hand, supervisors believed in establishing a high degree of operating structure in their dealings with employees and allowing them comparatively little control over their jobs (Stanton, 1959).

The data supplied by this investigation suggest that a company can offer an authoritarian approach to leadership and still show consideration for the individual employee. Unlike the negative implication suggested by the treatment of the individual under a totalitarian national state, this study failed to support the assumption on the part of some persons that an authoritarian approach in an industrial situation must, of necessity, be accompanied by a lack of consideration for the fundamental rights of the individual employee. Should further studies confirm these findings, it is quite possible that many of the negative connotations implied by the term "authoritarian" are unjustified when applied to industry. The results of this investigation would suggest that an organization may select a democratic, authoritarian, or a compromise approach to management and still be considerate of the feelings and desires of its employees.

The type of approach that a company's management may find most effective for its particular organization appears to be closely related to situational factors. It may very well be that, under certain circumstances, a democratic approach to management may be more effective, while in another situation, an authoritarian approach is preferable. Quite possibly, in a military situation, in a crisis, or when an emergency arises, firm, authoritarian, and directive leadership may not only be the most effective type of leadership, but also the style that is most welcome to all concerned. Another situational factor that must be considered in selecting a management approach is the particular leadership style preferred by the employees and the pattern to which they have become accustomed. It is suggested that a more democratic approach can be instituted only where employees are interested in and ready to assume responsibility for participation, identify with the goals and objectives of the organization, and have sufficient experience and knowledge to deal with the problems at hand. Where these qualifications are lacking or appear in limited degree, a more authoritarian or compromise approach may very well be most effective.

REFERENCES

BAVELAS, A., & LEWIN, K. Training in democratic leadership. *J. abnorm. soc. Psychol.*, 1942, **37**, 115–119.

BRADFORD, L. P., & LIPPIT, R. Building a democratic work group. *Personnel*, 1945, **22**, 142–152.

FLEISHMAN, E. A., HARRIS, E. F., & BURTT, H. E. *Leadership and supervision in industry.* Bur. of Educ. Res., 1955, No. 33. Columbus: Ohio State Univer.

FLEISHMAN, E. A. The leadership opinion questionnaire. In R. M. Stogdill & A. E. Coons (Eds.), *Leader behavior: Its description and measurement.* Res. Monogr., No. 88. Columbus: Ohio State Univer., 1957, 120–133.

GORDON, T. *Group-centered leadership.* Boston: Houghton Mifflin, 1955.

MCMURRY, R. N. The case for benevolent autocracy. *Harv. bus. Rev.*, 1958, **36**, 82–90.

NELSON, C. W. Development and evaluation of a leadership attitude scale for foremen. Unpublished doctoral dissertation, Univer. of Chicago, 1949.

NELSON, C. W. *The supervisory practices inventory.* Park Forest, Ill.: Mgmt. Res. Assoc., 1955.

STANTON, E. S. Company policies and supervisors' attitudes toward supervision. Unpublished doctoral dissertation, Columbia Univer., 1959.

TANNENBAUM, R., & SCHMIDT, W. H. *The dilemma of modern leadership: To be or not to be democratic.* Unpublished manuscript. Los Angeles: Univer. of Calif., 1957.

VITELES, M. S. *Motivation and morale in industry.* New York: Norton, 1953.

(Received March 19, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

A COMPARATIVE STUDY OF "TRANSFER THROUGH GENERALIZATION" AND "TRANSFER THROUGH IDENTICAL ELEMENTS" IN TECHNICAL TRAINING[1]

ARTHUR I. SIEGEL, MILTON RICHLIN, AND PHILIP FEDERMAN

Applied Psychological Services, Wayne, Pennsylvania

Recently, naval aviation technical training has shifted its focus from a general training program to a training program which introduces trainees directly into specialized instruction. It is the goal of this specialized training to produce men who are immediately useful upon graduation in a short "pipeline" time.

The newer program may be contrasted with the technical training previously given. This previous training was broader in nature, emphasized more deeply the theoretical aspects of the technical skills involved in maintenance, and relied to a greater extent on in-service training for imparting the specific technical skills needed for specific job performance. The purpose of the present study was to compare in a real situation the technical effectiveness of technicians given specific training (transfer through identical elements) with the efficiency of trainees given a more general background knowledge (transfer through generalization).

Graduates of each training program for three naval ratings were studied: (*a*) jet aviation machinist's mate, (*b*) air controlman, and (*c*) parachute rigger. Success as a parachute rigger or a machinist's mate depends upon mechanical ability and perceptual motor skill, while success as an air controlman depends to a considerable degree on verbal behavior and abstract reasoning.

Methods, Subjects, and Procedure

For each rating a complete library or listing of the tasks that the technician could be called upon to perform in the fleet was developed and cast in technical behavior check list (TBCL) form. Each check list contained three parts. Since it was felt that one of the most valid indicators of acceptable performance is the willingness of a man's supervisor to assign him without direct, technical supervision to various technical tasks, Part I determined the amount of time spent by the ratee on each of the tasks within the rating. In Part II, the amount of supervision that the striker required on each of the tasks he performed was requested. Part III acquired an estimation of the criticality, in terms of squadron mission, of each of the tasks listed. In order to study the developmental aspects of the ratees, separate evaluations were obtained in Parts I and II for each of two time periods: the first three months the man being rated was in the fleet (T-1) and the fourth to ninth months he was in the fleet (T-2).

The *Ss* were graduates of the two naval aviation training programs, the previous general "A" school program and the more recently established and specialized program. For the air controlmen 39 graduates of the more general training program (henceforth referred to as ACGs) and 42 graduates of the specialized program (ACTs) were studied on their fleet jobs.

For the parachute riggers, 10 graduates of the previous program (PRGs) and 23 graduates of the specialized training program (PRSs) were included. In the aviation motor machinist's mate rating, three groups were involved: graduates of the previous general training program (ADGs; $N = 21$); graduates who had received an intermediate type of training which involved specialized training but which did not include training on the specific equipment used in the fleet (ADJIs; $N = 60$); and graduates of the specialized program who had received specialized training and practice on the specific equipment found in the fleet (ADJSs; $N = 36$). The subjects were distributed over six naval air stations and 32 squadrons.

Derivation of Criterion Instrument

The mean time in hours spent during T-1 and T-2 by each group in each rating on each of the listed tasks was computed. For each task listed in Part II of the TBCLs, a score ranging from one to five was assigned: a score of five indicated proficient task performance after an initial checkout; a score of one was assigned for tasks on which the striker had received six or more checkouts but on which he was still unable to perform without direct supervision. Scores of four, three, or two

[1] This study was performed under Contract Nonr-2279(00) between Applied Psychological Services and the Office of Naval Research. We are indebted to D. Smith, John Nagay, and G. D. Mayo for their assistance throughout all phases of this work.

were respectively assigned as follows: proficient after 3 to 5 checkouts, proficient after 6 or more checkouts, 1 to 5 checkouts but not proficient. The mean and variance of the Part II scores for each task on each of the three separate TBCLs were computed by time period. Additionally, overall means and overall variances were computed for each task in each of the three separate TBCLs. Last, the criticality of each task was derived from the responses in Part III of the TBCLs.

Three item plots were then made, one for each of the ratings under consideration. The average variance of each task as computed from the TBCL, Part II, data was plotted as the ordinate value, while the abscissa point was determined by multiplying the mean time spent on the task by the criticality of the task as computed from the TBCL, Part III, data. In one sense and in measurement terminology, the ordinate represented the discriminating power of the items (tasks), while the abscissa value represented difficulty level. Inspection suggested fairly clear cutting points on each axis for eliminating items. Essentially, in developing the final criterion TBCLs, those tasks which were considered "unimportant" and which are relatively useless in a measurement instrument because of lack of discrimination (variability) between individuals were eliminated.

Thus, three separate final criterion TBCLs were developed, one for each of the ratings considered. All data henceforth reported are based on these final criterion TBCLs. The final criterion TBCL for motor machinists' mates contained 38 tasks; the final criterion TBCLs for parachute riggers and air controlmen contained 43 and 34 tasks respectively.

Results

Using the same scoring method discussed above, Part II of each TBCL was rescored with the total score being equal to the sum of the scored tasks divided by the number of tasks attempted. The total scores so derived were then subjected to a variance analysis. The results of these analyses are presented in Tables 1, 2, and 3.

These analyses enable a general answer to the question of whether the specialized training had exerted any general effects on fleet technical efficiency. For the aviation machinists' mates and parachute riggers no statistically significant between-groups differences were evidenced. For the air controlmen, the more generally trained group was superior to a statistically significant extent. The between-time period differences noted were to be expected since a trainee would be expected to perform at a superior level during his fourth to ninth months in the service as compared with his first three months.

TABLE 1

Analysis of Variance of Aviation Machinists' Mates Scores

Source of Variation	Sum of Squares	*df*	Var. Est.	*F*
Between cells	62,947	5	—	—
Between groups	16,697	2	8,349	1.65
Between time periods	39,737	1	39,737	7.84*
Groups × Times	6,513	2	3,256	$F < 1$
Within cells	905,288	178	5,086	—
Pooled error (4 + 5)	911,801	180	5,066	—
Total	968,234	183	—	—

* Significant at the .01 level of confidence.

The next step was to determine where the men in the air controlman's rating were doing well and where they needed improvement. The response score distributions for the five task content clusters included in the final criterion air controlman's TBCL were derived. These distributions are presented as Table 4. The distributions indicated that for each of the five content clusters (using equipment, using publications, testing equipment, receiving and transmitting messages, and controlling traffic), the generally trained group had higher

TABLE 2

Analysis of Variance of Air Controlmens' Scores

Source of Variation	Sum of Squares	*df*	Var. Est.	*F*
Between cells	87,103	3	—	—
Between groups	40,009	1	40,009	12.52*
Between time periods	45,333	1	45,333	14.19*
Groups × Times	1,761	1	1,761	$F < 1$
Within cells	448,751	140	3,205	
Pooled error (4 + 5)	450,512	141	3,195	
Total	535,854	143	—	

* Significant at the .01 level of confidence.

TABLE 3

ANALYSIS OF VARIANCE OF PARACHUTE RIGGERS' SCORES

Source of Variation	Sum of Squares	*df*	Var. Est.	*F*
Between cells	58,874	3	19,625	—
Between groups	7,897	1	7,897	1.91
Between time periods	49,385	1	49,385	11.97*
Groups × Times	1,592	1	1,592	$F < 1$
Within cells	217,021	52	4,173	—
Pooled error (4 + 5)	218,613	53	4,125	—
Total	275,895	55	—	

* Significant at .01 level of confidence.

means over both time periods. For the first time period both groups were poorest at the tasks involved in receiving and transmitting messages, while in the second time period the generally trained group was, by and large, proficient at these tasks, and the specifically trained group required additional training. Moreover, even in T-2, the specifically trained group remained weakest in receiving and transmitting messages and in controlling traffic. These tasks are believed to involve mostly nonroutine thinking as opposed to the specific information transfer involved in using publications, using equipment, and testing equipment.

Similarly, an examination of Table 4 from the point of view of within-group strengths and weaknesses during T-1 suggests that if publication use is ignored, the specifically trained group was best at equipment use and equipment testing, while the generally trained group was best at testing equipment and controlling traffic.

DISCUSSION

The usefulness of the present findings is dependent to a large extent on the criterion

TABLE 4

RESPONSE SCORE DISTRIBUTION FOR TOWER TASKS OF AIR CONTROLMEN

	T-1								T-2							
	Response %								Response %							
Tasks	1	2	3	4	5	*N*	*M*	σ^2	1	2	3	4	5	*N*	*M*	σ^2
Using equipment																
AC (G)	6	9	22	39	24	67	3.66	1.24	6	8	5	13	68	107	4.30	1.48
ACT	5	29	13	28	25	68	3.40	1.59	15	18	8	9	50	109	3.61	2.48
Using publications																
AC (G)	0	3	5	13	79	67	4.69	.48	0	0	1	3	96	90	4.94	.08
ACT	0	11	9	13	68	69	4.39	1.02	7	1	2	2	88	93	4.65	1.11
Testing equipment																
AC (G)	3	8	15	41	33	39	3.95	1.02	6	6	0	7	81	54	4.54	1.25
ACT	3	22	33	17	25	36	3.39	1.35	2	12	4	16	66	56	4.32	1.25
Receiving and transmitting messages																
AC (G)	0	14	18	35	33	88	3.53	1.04	2	10	2	10	76	124	4.50	1.07
ACT	9	39	21	4	27	94	3.01	1.84	15	19	7	10	49	129	3.61	2.48
Controlling traffic																
AC (G)	7	9	26	40	18	124	3.88	1.23	7	13	4	9	67	189	4.15	1.86
ACT	8	17	34	24	17	118	3.25	1.32	17	13	10	10	50	191	3.63	2.54
Totals																
AC (G)	4	8	19	34	35	385	3.88	1.20	4	9	3	9	75	564	4.42	1.35
ACT	5	24	23	17	31	385	3.44	1.66	13	13	7	9	58	578	3.85	2.30
Grand total	4	16	21	26	33	770	3.66	1.48	9	11	5	9	66	1142	4.13	1.92

used. A training criterion must possess certain characteristics if it is to be useful. These characteristics include relevance, comprehensiveness, consistency, correctly weighted elements, and analyzability. Taking these requirements in turn, the TBCL seems to have met the requisites of relevance and comprehensiveness. All of the included jobs were drawn directly from the training course and/or the actual job. Furthermore, any nontechnical work characteristics were purposefully omitted from the TBCLs. Consequently, the TBCLs were relevant by definition. Since space was allowed for addition by the rater of tasks not included in the list and since very few "write-ins" were made, it can be concluded that the technical jobs were comprehensively covered. Moreover, the fact that the TBCLs were designed as training criteria rather than as general fleet performance measurements instruments points to the necessity of excluding variance due to performance not associated with what was acquired in training.

The third characteristic, an analyzable criterion, has been amply demonstrated in the present study. Total TBCL scores were computed; the subsections were analyzed individually and combined; individual clusters of tasks were analyzed in terms of mean performance scores, variance, and response distributions. All of these analyses were possible in relation to time and subgroup differences.

The correct weighting of the elements of the criterion received indirect treatment in the present research. The item selection technique emphasized task criticality, frequency of occurrence, and variance. High variance tasks are inherently weighted more than low variance tasks. Moreover, the emphasis on criticality tended to eliminate routine jobs from the list. Thus, for instance, the AD TBCL included only 12 tasks involving routine line operations in comparison with 26 tasks involving major and intermediate inspections of jet aircraft and jet power plant removal and jet engine build-up.

In order to estimate the reliability of the TBCLs, product moment correlations were computed. In some cases a given striker was rated independently by two supervisors. In other cases, the same rater was asked to repeat his report. These data were combined in the reliability determination. It can be argued that the combined estimations are perhaps the best indication of total reliability since they consider both inter-rater and intra-rater reliability. The uncorrected reliability for the AD TBCL so determined was .84 for T-1 and .90 for T-2; for the AC and PR TBCLs the uncorrected reliabilities respectively by time period were .91, .89, .76, and .81.

In terms of the issue of general versus specific training, it seems that the specialized training for motor machinists and parachute riggers exerted few, if any, debilitating effects on the technical proficiency of the men involved. This contention does not seem to hold for the air controlmen. The differential effects of specialization may be explainable on the basis of the differences in skills involved in the various ratings. For the air controlmen, most of the tasks are of a conceptual nature, while for the parachute riggers and motor machinists mechanical manipulation is largely involved. Moreover, within the air controlman rating those tasks of a conceptual nature seemed to be the most difficult for the specifically trained air controlmen.

Summary and Conclusions

Technical Behavior Check Lists were used for assessing the fleet effectiveness of persons who had received two types of training: (1) specialized task oriented training, (2) general conceptual training. The effects of rapid, specialized, task-oriented training in the context with which we were concerned varied with the on-the-job performance requisites. For tasks involving mechanical manipulation few debilitating effects were introduced by the specialized training; for abstract conceptual tasks this contention does not seem to hold.

(Received March 23, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

THE VALIDITY OF BIOGRAPHICAL DATA ITEMS FOR FOOD COMPANY SALESMEN[1]

THOMAS W. HARRELL

Stanford University

A consulting company has advocated the screening of sales applicants on their background prior to the use of psychological tests. Based on experience a list of six "knock-out" factors has been developed. The consulting company has requested clients not to test an applicant for a selling job if one or more of these items is present. The list:

1. Instability of Residence

Regardless of the cause of changes of address (unless the applicant has been in the armed services or has been moved around by his employer), too frequent change of address points to lack of roots.

2. Failure in Business within Two Years

If your sales applicant failed in his own business, within two years prior to the time he applies for a sales job with you, he constitutes a risk.

3. Divorce or Separation within Two Years

A man who is undergoing active separation or divorce, or who was separated or divorced, even if remarried, within two years prior to his application, is an undue risk.

4. Excessive Personal Indebtedness

Some men will work hard when they have obligations to meet. However, when a sales applicant has excessive personal indebtedness (not including mortgage on home) he constitutes an undue risk. Indebtedness is considered excessive when it cannot be met within two years from earnings on the new job.

5. Too High a Previous Standard of Living

A Food Company made two items of this. One was previous income too high, which in this study meant a salary of $400 or more per month. The second was earnings less than living expenses.

6. Unexplained Gaps in the Employment Record

A study has been made of the validity of these background facts within the Food Company. The Company has used the "knock-out" factors during the past several years to exclude employment to sales applicants; although, due to the tight labor market and to the judgment of local Sales Supervisors, many men have been employed who have possessed one or more of the background facts. The Food Company has redefined No. 1, Insta-

[1] This study was sponsored by the Klein Institute for Aptitude Testing, Inc.

TABLE 1

DISTRIBUTION OF TYPE OF "KNOCK-OUT" FACTOR

Knock-out Factors	Combination Sales		Retail Sales					
			Successful		Resigned		Fired	
	N	%	*N*	%	*N*	%	*N*	%
Job-hopping	2	6	5	25	9	36	9	33
Previous income too high[a]	5	16	2	10	8	32	3	11
Earnings less than living expense	1	3	—	—	4	16	2	7
Business failure within past two years	—	—	—	—	2	8	1	4
Divorced within past two years	—	—	1	5	1	4	1	4
None	24	75	12	60	8	32	14	52
Number of cases	32		20		25		27	
"Knock-outs" per person	.25		.40		.96		.59	

[a] Previously earned salary of $400 or more set as "too high."

TABLE 2

STANDARD ERROR OF THE DIFFERENCE BETWEEN SUCCESSFUL AND UNSUCCESSFUL SALESMEN ON PRESENCE OF A "KNOCK-OUT" FACTOR

	Retail Salesmen					
	Successful		Resigned		Fired	
	z	Level of Significance	z	Level of Significance	z	Level of Significance
Combination sales (successful)	1.14[a]	—	3.26	1%	1.84	—
Retail sales (successful)			1.88	—	.54	—
Resigned retail (unsuccessful)					1.40	—

[a] Method used was standard error of the difference between two uncorrelated proportions. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950, Pp. 76–77.

bility of Residence, as Job-hopping. For the purpose of the recent study this means having had three or more jobs during the preceding five years, or two jobs in the past two years and obviously to be applying for another.

The problem was whether or not there was any difference in the background facts between a group of successful employees and a group of unsuccessful employees. The selection of persons for each of these samples was initially made for another purpose based on whether the man had been tested on a series of psychological tests. The successful employees were made up of two samples. 1. Twenty Retail Salesmen were chosen from approximately 200 employed. They were selected as having been employed for at least two years and having a high rating of competence by immediate Supervisor. The job of the Retail Salesmen is to promote the sales of canned goods through calls on grocery stores and through setting up displays. 2. Thirty-two Combination Salesmen were studied. Combination Salesmen were promoted from Retail Salesmen. Combination Salesmen sell to Jobbers as well as calling on grocery stores for the same purpose as Retail Salesmen. Unsuccessful employees from the company's point of view were 27 Retail Salesmen who had been fired and 25 who had resigned. The samples of Combination Salesmen and of unsuccessful Retail Salesmen who were fired or who resigned were not taken from the group of 200 Retail Salesmen then employed from whom the sample of 20 Successful Retail Salesmen was drawn.

The distribution of type of "knock-out" factors for each of these four groups is shown in Table 1. There were no cases of excessive personal indebtedness nor unexplained gaps in the employment record and hence No. 4 and No. 6 are not shown. The highest frequency, .96 "knock-outs" per person, was present in the salesmen who had resigned. In

TABLE 3

DISTRIBUTION OF KNOCK-OUT FACTORS

Knock-out Factors	Combination Sales		Retail Sales					
			Successful		Resigned		Fired	
	N	%	N	%	N	%	N	%
None	24	75	12	60	8	32	14	52
One	8	25	8	40	12	48	10	37
Two					3	12	3	11
Three					2	8		
Total	32		20		25		27	

TABLE 4

STANDARD ERROR OF THE DIFFERENCE BETWEEN SUCCESSFUL AND UNSUCCESSFUL SALESMEN ON NUMBER OF "KNOCK-OUT" FACTORS

	Retail Salesmen					
	Successful		Resigned		Fired	
	z	Level of Significance	z	Level of Significance	z	Level of Significance
Combination sales (successful)	1.14	—	4.00	1%	2.26	5%
Retail sales (successful)			2.52	5%	.90	—
Resigned retail (unsuccessful)					1.48	—

order of frequency the other groups follow as Fired Salesmen, .59, Successful Retail Salesmen, .40, and Combination Salesmen, .25. Job-hopping is the most frequent of the factors present.

The significance of the differences between the number of successful and unsuccessful salesmen who had at least one "knock-out" factor is shown in Table 2. Salesmen who resigned were different from Combination Salesmen at the .01 level but were not different from Successful Retail Salesmen. Fired Salesmen did not differ from Combination Salesmen nor Successful Retail Salesmen with respect to presence of a "knock-out" factor.

The distribution of number of "knock-out" factors is shown in Table 3. No successful Retail or Combination Salesman had more than one "knock-out" factor while five of the Resigned Salesmen and three of the Fired Salesmen had two or more. Using the total number of "knock-out" factors shown in Table 1, the significance of the differences between Successful and Unsuccessful Salesmen is shown in Table 4. Salesmen who resigned had significantly more "knock-out" factors than the Combination Salesmen at the .01 level and than the Retail Salesmen at the .05 level. Fired Salesmen had significantly more "knock-out" factors than Combination Salesmen at the .05 level, but did not have significantly more "knock-out" factors than the Retail Salesmen.

SUMMARY AND CONCLUSIONS

1. The presence of a "knock-out" factor was not significantly different between a sample of Successful Retail Salesmen and each of two groups of Unsuccessful Retail Salesmen who were fired and resigned.

2. Combination Salesmen who had been promoted from Retail Salesmen, were less likely to have "knock-out" factors than Resigned Salesmen, as shown by a difference at the .01 level.

3. The number of "knock-out" factors was greater among Resigned Salesmen than Successful Retail Salesmen at the .05 level. The number of "knock-out" factors was greater for Resigned Retail Salesmen than for Combination Salesmen at the .01 level. The number of "knock-out" factors was greater for Fired Retail Salesmen than for Combination Salesmen at the .05 level.

4. The most frequent "knock-out" factor was job-hopping.

5. The results confirm the value of some use of the "knock-out" factors, although the labor market may not permit being as rigorous as the consultant recommended. There is an indication that the presence of two of the factors is a better cutoff than the presence of only one. The number of cases is too small to examine the significance of having two or more "knock-out" factors.

(Received April 15, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

THE EFFECTS OF TASK COMPLEXITY AND TIME PRESSURE UPON TEAM PRODUCTIVITY[1]

PAULINE N. PEPINSKY AND HAROLD B. PEPINSKY

Ohio State University

AND WILLIAM B. PAVLIK

Western Michigan University

This experiment simulated an organization in which a team of workers, whose actions are structured by management, performs a series of repetitive, coordinated, and interlocking operations. Under such conditions, is the productivity of the team significantly affected by the amount of time pressure under which the team works? If so, does variation in time pressure have differential performance effects dependent upon the complexity of the team's task? In early research on performance in industrial settings, emphasis was placed upon job specifications, time and motion study, and the effects of special working conditions. Then studies of behavior in organizations shifted to problems of interpersonal relations and group process. But now there is evident a revived interest in examining the physical characteristics and effects of the stimulus situation.

This "new look," however, occurs against a different conceptual ground. "Working conditions" are viewed as one set of variables that, in interaction with personal, social, and economic antecedents, modify group interaction and in turn may affect tangible output on an allocated team task. In the present study a deliberate effort was made to control the effects of differences in individual abilities, group structure, and organization in order to assess in a precise way the effects of the particular situational variables of task complexity and time pressure. Information about the effects of task complexity and time pressure also is of interest in relation to diverse theoretical attempts to account for individual behavior associated with strong response competition. Recurring empirical evidence indicates that decrement in individual performance occurs under special conditions of increased pressure or deprivation, where increases in drive level can be inferred as well (Taylor, 1956).

Given two team tasks, one (Task S) relatively simple, and the other (Task C) relatively complex; and three time conditions, designated as low (L), medium (M), or high (H) time pressure, the central hypothesis was that there would be a significant interaction between the effects upon team productivity of Tasks S and C and Time Cond. L, M, and H. Specifically, it was predicted that comparisons among the means involved in this interaction would show:

(*a*) Team productivity on Task S to be significantly greater under Cond. H than under Cond. L;

(*b*) Team productivity on Task S under Cond. M to be intermediate between productivity under L and H;

(*c*) Team productivity on Task C under Cond. M to be significantly greater than productivity under either L or H.

METHOD

Subjects. Seventy-two volunteer male *Ss* were assigned to 24 three-man teams. The *Ss* were enrolled in the university introductory psychology course. The only other restrictions in selection were the exclusion of individuals whose obvious physical handicaps would have prevented their performing an assembly task, and the requirement that each team be comprised of previously unacquainted persons.

Experimental Tasks. Task complexity was established at two levels. Each task involved the repeated assembly and disassembly by a team of as many

[1] This research was done under Contract n6ori-17, T.O.111(NR 171-123) with the Office of Naval Research, but the views expressed are those of the authors alone. This paper is a condensation of a more detailed technical report (Pepinsky, Pepinsky, & Pavlik, 1956), which may be obtained on loan from the Gifts and Exchange Department of Ohio State University Library. The authors wish to acknowledge the valuable statistical advice of D. Ransom Whitney.

replicas of a nonrepresentational Tinkertoy model as possible during a 20-min. session. The possible effect of organizational differences was controlled by the stipulation that all teams follow a rehearsed procedure, performing interlocked operations in a specified order, with the role and function of each *S* allocated by *E*. Between sessions the position of *S*s were rotated so that each worked with a different type of component on each task.

Task complexity was defined by (*a*) the number of operations required to assemble the model, (*b*) the necessity for simultaneous coordination among the team members to perform some operations—in contrast to operations performed in sequence by individual *S*s, (*c*) the amount of variety in the ordered pattern of assembly operations, and (*d*) the number of spatial dimensions in the completed model. The physical arrangements and *E*'s procedures were the same for both the S and C tasks, except for adjustments to correspond to the variation in complexity.

Measure of the Dependent Variable. A team's score on each task was an adjusted total of the number of operations performed during a work session. The exact procedure used to determine the scores is specified in Pepinsky, Pepinsky, and Pavlik (1956, pp. 84–85). The numbers of operations credited for disassembly of both models and for final assembly of the complex model were determined by the mean proportionate amounts of time spent by all 24 teams in these particular phases of the tasks. A ratio based on all teams was considered to be a more reliable weighting than one computed for each team separately. The only distortion that the weighting system could introduce seemed to be a possible overall reduction in the variability of the scores, with the concomitant risk of conservative error in the final conclusions.

Procedure. For each of the experimental teams the procedure consisted of the same consecutive phases: (*a*) a preliminary session of supervised practice on the first task, (*b*) the first 20-min. experimental session, (*c*) a 5-min. recess, (*d*) a session of practice on the second task, (*e*) the second 20-min. experimental session, (*f*) administration of a postexperimental questionnaire, and (*g*) a brief decompression session. All sessions were held at the same time of day.

For each task the team built successively a specified number of models, which constituted roughly comparable work units (a set of six simple models or a single complex model). The team had each unit checked by *E*, and then disassembled these models before beginning new ones. To prevent attenuation of the effect of time pressure through tension release in casual conversation and joking, *E* stated that team members were permitted to discuss the task only. During the experiment, *E* did not interact with *S*s, except to indicate by a nod or a gesture that completed models were correctly assembled or should be corrected.

The three time conditions were manipulated by varying the frequency of standard verbal signals given over a loudspeaker. Condition L consisted in one signal given at the end of 10 min., Cond. M, a signal every three min. up to 18 min., after which the stop signal followed in two minutes; Cond. H, signals delivered at an accelerated pace beginning with one every minute and increasing to one every 15 sec. The signals consisted simply in the timekeeper's saying, "_____ minutes to go," or toward the end of the session, "_____ minutes," or "_____ seconds."

Experimental Design. The experiment was conducted according to a factorial design, counterbalanced to permit separate examination of the effects of the time conditions, the tasks, time condition sequences, task sequences, and blocks. In one block of 12 groups each task, placed once in first and once in second order, occurred in combination with each time condition, placed in each of the six possible sequences of three conditions, taken two at a time. The design included two blocks or one complete replication. All of the independent comparisons were allotted to the effects of these variables and their interactions.

Results

Checks upon the Experimental Conditions. The *S*s' responses to five postexperimental questionnaire items tell us (*a*) whether they identified Task C as more complex than Task S, and (*b*) whether their discriminations among the time conditions were consistent with the experimental attempt to vary time pressure. All but four *S*s identified Task C as the more complex. A significant proportion (p less than .01) liked working on Task C better than Task S and preferred the second session to the first, but *S*s' preference for one session over the other was not related to the time conditions. Regardless of either task or time conditions, *S*s in general tended to believe (p less than .001) that they were more productive during the second session.

A significant majority of *S*s (p less than .001) did feel under more pressure and were more aware of the passage of time during the session in which they worked under the higher of two time conditions (i.e., either H as opposed to M, or M as opposed to L). There were tendencies (p less than .10 but greater than .05) for *S*s' feelings of being under pressure to occur with relatively greater frequency during work on Task C and during the second work session. There is no evidence, however, that task or session alone had any relationship to *S*s' reported awareness of the passing of time.

TABLE 1

MEAN PRODUCTIVITY SCORES ON TWO TASKS UNDER THREE TIME CONDITIONS

Task	Time Condition L	M	H	Grand Mean
S	590.05	615.21	611.78	605.68
C	715.61	728.17	675.50	705.43
Grand mean	652.83	671.69	643.64	

Performance during Separate Work Sessions. Table 1 presents the sets of mean scores most directly relevant to the hypotheses. By inspection, the Task C scores are consistently larger than the Task S scores, and there seems to be a curvilinear variation in productivity from Conditions L to M to H. The significance of these apparent trends was tested first by means of a complete analysis of variance of single session scores (shown in Pepinsky, Pepinsky, & Pavlik, 1956). This basic analysis did not support the central hypothesis. It demonstrated unequivocally that differences between tasks and the interaction of tasks and task sequence are primarily responsible for the variability in total team scores in separate work sessions: scores are significantly higher on Task C than on Task S, and higher in the second session than the first.[2]

Performance over Time. As an alternative to the analysis of performance in separate sessions, *difference* scores were examined to determine whether changes in the productivity of the *same* teams from one session to the next reflected a curvilinear relationship to changes in time pressure. In order to obtain a basis of reference, the 24 *first* session scores were analyzed separately. This preliminary analysis of variance (not shown) shows that the main effect of tasks (with Task C producing larger scores) is significant at the .01 level and substantially accounts for the variance in these "first" measurements. The Time Conditions do not have a significant effect upon the first session scores, nor do they interact significantly with the Tasks.

As shown in Table 2, an analysis was made of changes observed in the second session, when the same teams performed on different tasks under different time conditions. The measures employed are the 24 algebraic differences obtained by subtracting the Session 1 scores from the Session 2 scores. The three independent variables involved here are defined as follows:

(*a*) Two Changes in Task—S → C, or C → S.

(*b*) Three Time Condition Pairs—L ⇆ H; L ⇆ M; or M ⇆ H, taken without regard to order of presentation, so that the Class L ⇆ H includes all teams that performed under L in one session and H in the other, regardless of which came first and which second.

(*c*) Two Directions of Change in Time Condition—"higher" time condition in the second session than in the first, or "lower"

TABLE 2

ANALYSIS OF VARIANCE OF DIFFERENCES BETWEEN PRODUCTIVITY SCORES OF SESSION 1 AND SESSION 2 FOR 24 TEAMS

Source of Variation	*df*	Mean Square	*F*
Change in task	1	243,587.305	186.57
Time condition pair	2	1,112.816	—
Direction of change in time condition	1	642.631	—
C Ta × Ti C P	2	2,264.230	1.73
C Ta × DC Ti C	1	782.842	—
Ti C P × D C Ti C	2	9,427.995	7.22*
C Ta × Ti C P × D C Ti C	2	1,286.036	—
Error	12	1,305.608	
Total	23		

* Significant at .01 level.

[2] The apparent task difference cannot be attributed to the extraneous effect of the amount of time consumed by *E* in the checking operation. A *t* test of the difference between the two tasks in mean operations per second of available work time shows that a reliable difference in favor of the complex task remains (p less than .001). Differences among mean operations per second under the three time conditions are not statistically significant, although the mean production rates for both tasks under Condition M appear to be slightly higher than under L or H. Plotted curves (not shown) of the median times spent by the 24 teams on assembly and disassembly of successive models show within-session trends consistent with the experimental predictions. But these tendencies occur within too restricted a range to have a significant effect upon total scores.

time condition in the second session than in the first, i.e., increase vs. decrease in time pressure.

All of the subclass means are positive in sign (i.e., indicating that Session 2 scores show increases over Session 1) except for the change in task from C to S. Comparisons between tasks consistently show a pronounced discrepancy in favor of C. Since Task C scores are higher than Task S scores, and Session 2 scores are higher than those of Session 1, the difference between C and S is increased when C occurs in the second session. Consequently, Change in Task, treated as a main effect, exerts an overwhelming influence in comparison to either of the two Time Condition variables, and it is clearly independent of both.

What emerges in Table 2 as a new finding is the significant (p less than .01) first-order interaction between Time Condition Pair and Direction of Change in Time Condition. (Application of Bartlett's test indicates that the variances of the six subclasses involved in this interaction are homogeneous.) While Time Conditions have no observable effect upon the first session scores, changes in the second session scores are differentially affected by Directions of Change in Time Condition occurring in different Time Condition Pairs. Thus, if Condition L is followed by H, a smaller increase occurs than if L is followed by M. In the instance of the L ⇆ H pair, a shift in the L to H direction produces a smaller increase than a change from H to L. These trends are reversed in the case of the L ⇆ M combinations, i.e., L to M produces a larger increase than M to L. The discrepancies associated with direction of change in the M ⇆ H pair have relatively small values.

When task (a mean difference of about 100 points) and practice (a mean difference of about 64 points) effects are removed, it becomes apparent that adjusted score differences ascribable to changes in time pressure alone are *not* consistently positive in sign. Irrespective of task, a gain in productivity accompanies an increase in pressure from L to M, but a decrement occurs when pressure increases from L to H; the latter change in performance is larger in magnitude than that associated with a reduction in pressure from H to L. Although a curvilinear relationship of time pressure to team performance was predicted for Task C only, it turns out that over an extended period, when *changes* in pressure upon the same teams occur, this effect characterizes their performance on both tasks.

Discussion

The most striking finding is the predominant effect upon team productivity of the difference between the simple and the complex tasks; greater task complexity is consistently associated with greater productivity. An increase in task complexity does have a positive motivational effect through maintaining interest in the performance of repetitive operations. This interpretation is supported by the statements of *Ss* independent of their measured team performance.

Although *Ss* reported that they felt under more pressure and were more aware of the passage of time when the time signals were increased in frequency, their responses to more or less frequent signals were a matter of indifference in terms of perceived reward value. Incentive properties are ascribed to time pressure conditions as they are clearly made to connote present or potential success or failure in a given situation; otherwise, interposed time signals may be little more than a trivial distraction.

Despite the initial attempt to minimize the effects of learning, the second session scores still show a practice effect, which is independent of the task and task sequence differences. Since the effect of practice is virtually the same no matter whether Task S or C is presented first, the two tasks evidently are not widely discrepant in difficulty. Both appear to lie within what Day (1956) has called an "optimal zone of performance," although they are discriminably different in respect to stimulus complexity. Failure to corroborate the central hypothesis in respect to single session scores may have been a function of comparability in task difficulty. But an increase in this task differential also would have led to a differential contamination of performance with learning on the more complex task.

It may be that even if it is highly prestructured, the team (as contrasted with the individual) work situation is indigenously so complicated that over an extended period high

time pressure is apt to produce a decline in productivity, even on an apparently quite simple group task. This inference is compatible with the finding that *differences between sessions* in productivity show a significant and curvilinear relationship to *changes* in time pressure, regardless of changes in task. In individual task performance accuracy under high pressure may vary as a function of initial differences in drive level. But a team engaged in tightly interlocked operations may be no more competent than its "weakest link," i.e., the member most predisposed toward loss of efficiency as pressure increases beyond a certain point.

The experiment is a beginning attempt to deal with a complicated and important problem. Present results underscore the importance of specifying both the operational and the phenomenal stimulus characteristics of tasks employed in research on behavior in groups and organizations and the likelihood that task variables may account for apparently contradictory results in otherwise similar studies. This study also provides tests of the effects upon team performance of two situational variables that previously have been used in conjunction with other procedures (cf. Lanzetta, 1955; Schutz, 1955) without precise knowledge of their independent predictive value.

The experimental findings have several practical implications for situations where small, trained teams are employed in the performance of highly structured, repetitive, assembly tasks. It seems probable that such teams will be more motivated to increase output, if their assigned tasks are sufficiently complex to provide some intrinsic interest and variety than if they are so simple in terms of the workers' ability as to be highly monotonous and boring. Whether the task is very simple or more complex, an acceleration in time pressure up to a moderate level can produce an acceleration in output, but an increase from low to high pressure is apt to have an adverse effect.

Summary

An experiment was designed to test the effects of the complexity of repetitive group tasks and of varied conditions of time pressure upon the output of structured work teams. Twenty-four three-man teams each performed two practiced assembly tasks, one relatively simple, and the other relatively complex. Time pressure was established at three levels by varying the frequency of interspersed oral signals, which announced the time remaining in a given work session. Team productivity was defined as an adjusted total of the number of operations performed on each task.

The central hypothesis that there would be a significant interaction between task complexity and time pressure is not supported. There is, however, positive evidence (*a*) that the teams were more productive when working on a task that was sufficiently complex to reduce boredom, but (*b*) that, irrespective of task, when changes in time pressure occurred, the effect upon changes in team productivity was curvilinear.

REFERENCES

Day, R. H. Relative task difficulty and transfer of training in skilled performance. *Psychol. Bull.*, 1956, **53**, 160–168.

Lanzetta, J. T. Group behavior under stress. *Hum. Rel.*, 1955, **8**, 29–52.

Pepinsky, Pauline N., Pepinsky, H. B., & Pavlik, W. B. *Motivational factors in individual and group productivity: III. The effects of task complexity and time pressure upon team productivity.* Columbus: Ohio State Univer. Res. Found., 1956.

Schutz, W. C. What makes groups productive? *Hum. Rel.*, 1955, **8**, 429–465.

Taylor, Janet A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

(Received April 16, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

AN INEXPENSIVE METHOD TO DETERMINE THE EFFICIENCY OF A TELEVISION PROGRAM[1]

LAWRENCE MYERS, JR. AND ERIC F. GARDNER

Syracuse University

Constantly rising television costs occasioned by expanding program costs and an expanding list of markets into which programs are broadcast make the rapid determination of audience information increasingly desirable to the television advertiser. The television rating services, while providing a considerable amount of audience information for the major markets, provide infrequent reports or none at all for many minor markets. Information relative to these smaller markets is nevertheless important to the advertiser who is purchasing time on stations in these areas. Further, no information is provided for any market on a weekly basis. It would appear that a need exists for the development of a rapid, valid, and inexpensive technique to assess the commercial effectiveness of a television program in each or all of the markets into which it may be broadcast.

PROBLEM

A number of quantitative approaches are available to evaluate a sponsored television program. Certainly a useful criterion is sales resulting from a given expenditure of television advertising dollars, but only in unusual circumstances is such information, particularly on an individual program or market basis, ever available.

A second criterion used to infer television program effectiveness is an estimate of audience size. The need for, and usefulness of, program audience size measurements led to the development of the rating services. Stated in terms of ratings, the advertiser learns the percentage of households in a given area that is exposed to his television program and, presumably, to his advertisement.

A third criterion to determine the success of a television program is efficiency, an indication of the audience obtained per dollar invested. On a national basis, it has been regular practice for years for the advertiser, or his agency, having obtained an estimate of the size of the broadcast audience from a rating service, to compare the total audience with the total cost to determine the unit cost, or efficiency. Although one may find exceptions to the practice, efficiency normally is defined as the cost in dollars to reach 1,000 households ($/M).

The use of efficiency circumvents problems associated with audience size when evaluating various ratings for individual markets. One market may contain more homes than another. Identical ratings in both markets will not indicate identical audiences. The relationship is further complicated by the factor of cost. Even if two markets produce audiences of equal magnitude, the cost to reach one audience may be greater than the cost to reach the other. A satisfactory solution to this problem is to substitute unit costs, or efficiencies, as these are comparable from market to market.

It is recognized that once the audience tunes to a program many factors other than cost—for example, program image, mood, holding power, commercial identification, psychological appeal—become important in achieving maximum commercial effectiveness. Such problems must be approached and corrected on an individual basis and are outside the scope of this article. All other things being equal, efficiency, as indicated by the dollar cost to present the commercial message in each 1,000 homes, remains as a very useful criterion of television advertising success.

If television program efficiency, as defined, is a useful concept, can it be determined market-by-market, quickly, and more economically than program ratings are currently determined. One can estimate program efficiency, involving costs and audience size, on a market-by-market basis as satisfactorily as

[1] This paper represents part of a dissertation submitted by the senior author to Syracuse University in 1956 in partial fulfillment of the requirements for the PhD degree.

one can estimate audience size measurements alone. If enough interviewers can be employed coincidentally with or shortly after a program and the set usage data can be rapidly processed, then the problem of speed is solved. Economy remains as the major problem, and it seems clear that a significant economy will be effected only by a decrease in the size of the sample. Smaller samples may decrease survey costs to the point where more frequent measurements in more cities may be possible.

The purpose of this paper is to describe and demonstrate how the technique of sequential analysis, applied to television audience size information obtained from a random sample, can be utilized to produce useful estimates of the efficiency of a television program quickly and economically.

Method

To review briefly the theory developed by Wald and others (Statistical Research Group, 1945; Wald: 1945, 1947), any sequential plan consists of examining, in sequence, items chosen at random from a lot and making, with stated risks of error, one of three decisions: (*a*) that the lot is acceptable, (*b*) that the lot is unacceptable, (*c*) that the cumulated evidence is insufficient for either decision without too great a risk of error. If the third decision is made, an additional item is inspected and the same three decisions are reconsidered on the basis of the additional information supplied by the new item. The inspection normally continues until the data are sufficient to warrant the acceptance or rejection decision.

In general, the two lot qualities, or two hypotheses, H_1 and H_2, between which one is trying to distinguish, are selected so that each of them and their difference is practically meaningful. Likewise, the risks, α of rejecting H_1 when H_1 is true and β of accepting H_1 when H_2 is true, are chosen with an eye to the practical effects of each kind of error.

The sequential design requires that a person set up certain "rules of the game" before he plays the game. First, he must decide upon the two hypotheses which state the acceptable and unacceptable efficiency limits with which he is primarily concerned. For example, an advertiser may stipulate that he hopes his program will achieve an efficiency of \$10/M, but will be displeased if his program achieves an efficiency worse than \$20/M. That is, an advertiser may very much desire to reach homes with his commercial message at a cost of one cent per home or less; but he will be rather unhappy if his cost per home exceeds two cents. He is less concerned with intermediate values. Also, the efficiency of major network evening television programs approximates \$10/M. One may have very good reasons for selecting other efficiencies for H_1 and H_2. For example, in a large and costly market an advertiser might require an efficiency of \$7.50/M and be unhappy with an efficiency below \$15/M; while this same advertiser, in order to achieve as near complete saturation as possible, might accept \$15/M in some smaller markets and be unhappy only if his efficiency fell below \$30/M.

In addition to selecting two meaningful efficiencies, one must announce two other "rules of the game." First, he must decide upon the value of α, the maximum risk or probability of rejecting an efficiency denoted by H_1. In the sequential plan, a rejection of H_1 automatically implies an acceptance of H_2. Second, he must decide upon the value of β, the maximum risk or probability of accepting H_1 when H_2 is true. One must assign values on the basis of meaningful risks to the user. For example, one would normally wish that the chances of rejecting a satisfactory efficiency would be low, say perhaps one in twenty. Conversely, one might not be quite as concerned about the possibility of accepting an unsatisfactory efficiency, and so permit a risk of perhaps one in ten that this type of incorrect decision might be made. Although a risk always exists, therefore, that a satisfactory television program efficiency will be erroneously rejected or that an unsatisfactory television program efficiency will be erroneously accepted on the basis of evidence from a sample, one is privileged to establish the limits of risk prior to the survey.

Those familiar with the use of sequential analysis in most inspection processes will recall that because the majority of items inspected will meet prescribed specifications one is normally concerned with the rapid determination of the percentage of defective items. However, when conducting a survey to determine the proportion of homes exposed to a television program, the logical method is to record successes, or homes exposed to the program.

Having established desirable and undesirable efficiencies for a certain television program as well as risks of making incorrect decisions on the basis of the sample evidence, and knowing the costs and the number of homes in a market, a Sequential Decision Table may be constructed. The table will indicate, for successive sample sizes, the number of homes exposed to the program—denoted by V_1 and V_2 for "viewer"—associated with the two hypotheses, H_1 and H_2; and will include a blank column upon which one may record and cumulate the actual number of home exposures determined from the interviewing.

Sequential analysis, as a procedure to examine sample items sequentially and to cumulate the evidence until a decision is reached, assumes that the characteristics of the sample, as well as the universe from which the sample is drawn, remain constant during the period of the inspection process. Because the numerical size of audience exposed to a television program may fluctuate during the program, the survey method should produce a "total audience" rather than an "average audience" statistic. Such data may be obtained by inquiring of a sample immediately following the program (or program segment) regarding set usage during the program. At this time the total audience, that is, the number of homes ex-

posed to all or part of the program, is completely established, and one may with confidence proceed to estimate that size sequentially or otherwise.

The telephone interview (immediate recall) method is suggested as a satisfactory method to determine total audience. If all calls concerned with a half-hour program are completed within 30 min. of the program's conclusion, a sufficiently short length of time will elapse prior to a respondent being asked to recall television set usage to preclude significant memory lapse. Myers (1956) has demonstrated that, for television program audience size estimation, a telephone sample is sufficiently representative of all homes in a market in which residential telephone subscriptions are 80% or higher. Use of a properly selected telephone probability sample also satisfies the sequential requirement that each interview be drawn at random from the population and be independent of all other interviews.

Using a previously selected sample of homes, a small team of telephone interviewers conducts calls from a central location, thus permitting consecutive accumulation and analysis of the sample data according to sequential requirements. As set usage data is obtained by each operator, an inspector keeps a cumulative account of the number of homes reached and the number of homes tuned to the program in question. By comparing these running totals with an appropriate sequential decision table, a decision can be made within minimum time and with a minimum sample.

An Illustration

To illustrate the procedure, assume that it is desired to assess the efficiency of the "Ford Theater" television program in the city of Syracuse, New York. The advertiser establishes efficiencies of \$10/M (desirable) and \$20/M (undesirable), as well as risks of $\alpha = .05$ and $\beta = .10$ of making incorrect decisions on the basis of the sample evidence. (Let us assume that, associated with these efficiencies for this market, the program costs —time charges and prorated production costs —are \$900 and the number of homes is 300,000.) It should be re-emphasized that these specifications of efficiencies and risks completely determine the sequential sampling plan and permit one to set up tables to facilitate the collection and evaluation of subsequent survey data. Table 1 shows the Sequential Decision Table conforming to the above specifications. The table also shows the results of an actual series of interviews to ascertain the efficiency of "Ford Theater."

Examination of the table shows that the first home interviewed saw the program, as did the second home. The third through the thirteenth homes did not see the program for one of a number of reasons. The fourteenth, sixteenth, nineteenth, and twenty-first homes were tuned to the program. If one halts the interviewing momentarily and examines the data accumulated at the conclusion of the twenty-first interview, one notes that the number of homes found viewing the program ($V_{cum} = 6$) is larger than V_2 (1) so that one cannot make the decision that the efficiency is poorer than \$20/M. At the same time, V_{cum} is smaller than V_1 (8) so that one is likewise unable to make the decision that the efficiency is better than \$10/M. Interviewing is therefore continued.

TABLE 1

Sequential Decision Table for "Ford Theater"

(H_1 = \$10/M H_2 = \$20/M $\alpha = .05$ $\beta = .10$)

N	V	V_1	V_{cum}	V_2
1	x	—	1	—
2	x	—	2	—
3		—	2	—
4		—	2	—
5		4	2	—
6		4	2	—
7		5	2	—
8		5	2	—
9		5	2	—
10		5	2	—
11		6	2	—
12		6	2	—
13		6	2	—
14	x	6	3	—
15		6	3	0
16	x	7	4	0
17		7	4	0
18		7	4	0
19	x	7	5	0
20		8	5	1
21	x	8	6	1
22		8	6	1
23	x	8	7	1
24		8	7	2
25		9	7	2
26		9	7	2
27	x	9	8	2
28		9	8	3
29[a]	x	9	9	3

Note.—Interviewing is concluded after the 29th completed call, and the hypothesis, H_1, that the efficiency of "Ford Theater" was better than \$10/M, is accepted.

The twenty-third, twenty-seventh, and twenty-ninth homes are found to have viewed the program. At this point the cumulated homes that viewed the program equaled the V_1 acceptance number of 9, so the interviewing was terminated with the decision that the efficiency of "Ford Theater" was better than $10/M.

The efficiency of the program based on an estimate from a larger sample size of 267 was later computed to be $8.50/M, verifying the earlier decision made after a sample of only 29 homes. Numerous examples demonstrating the conformity of empirical data to theory may be found in Myers' dissertation (1956).

As will be noted from the illustration, a unique feature of any sequential analysis plan is the fact that the size of the sample is unknown prior to the survey. Practically, it will be necessary to decide arbitrarily on a termination point for the interviewing procedure. Although sample size is undetermined prior to the survey, it is possible to compute the sample size that will be required, on the average, for decisions when repeated surveys are made with a given sequential plan. The average number of interviews will depend upon the characteristics of the plan and upon the efficiency of the program being surveyed. Specifically, the smaller the risks, the greater the amount of inspection; also the smaller the difference between the two efficiencies, the greater the amount of inspection; and the greatest average amount of inspection is required for those programs whose efficiencies lie between the two hypotheses. An Average Sample Number (ASN) curve (Statistical Research Group, 1945) associated with a sequential plan can be drawn to show the relationship between various program efficiencies and the average sample size required for each. The maximum ordinate of the ASN curve—denoted by $\tilde{n}_s$—will indicate the program efficiency that will require the greatest average amount of inspection. Since the most unfavorable condition for the research in terms of the amount of sampling needed is given by the maximum ordinate, it is recommended that the average sample size represented by $\tilde{n}_s$ be established as a satisfactory termination point for the survey.

In most instances decisions will be made prior to reaching the recommended sample size, as was the case with the "Ford Theater" illustration where the decision was made at $n = 29$ when $\tilde{n}_s$ was actually computed to be 48. In the event, however, that it is impossible to reach a decision prior to interviewing a sample of size $\tilde{n}_s$, then the survey should be concluded with no decision but with the implication that the cost per thousand probably lies somewhere between the two designated limits.

Discussion

Does this application of sequential analysis to the telephone survey method effect a substantial savings in cost? The major cost item involved, once the sequential decision plan is prepared, is that of interviews. Using efficiency limits and risks described above, most television programs that are broadcast during the evening hours will need only two telephone operators to obtain a sample sufficiently large to reach a decision. If most efficiency decisions can be made with two operators—with a few programs requiring three or four operators at the most—the opportunity to obtain similar data frequently in many markets becomes economically feasible.

The sequential model will be of primary concern to advertisers who use either network or spot campaigns to sponsor television programs in many markets. To ascertain general strengths and weaknesses quickly, these advertisers should be interested, among other things, in a continuous running account of their program efficiencies in each market; and they may, therefore, adopt sequential models that satisfy their advertising and marketing plans if such data is obtainable for only a few dollars per market. One will note that this method deliberately precludes any mention of program ratings with their accompanying measurement errors. Likewise, the method does not yield the actual efficiency obtained by a television program. Rather, a definitive efficiency concept has been employed; namely, that for each market, a television program's efficiency is better than a designated value, or worse than another designated value, or probably lies somewhere between the two.

An advertiser thus has available an inexpensive and rapid technique for evaluating

the efficiency of a television program on a market-by-market basis. Whether he presents his program in 20, 50, or 150 markets, he can maintain weekly checks on his efficiency in each market at a fraction of the cost normally required to produce individual program ratings. As weaknesses are exposed in various markets, appropriate individual analysis and action can then be taken to correct these deficiencies.

Summary

This paper has presented a new concept and method for determining the commercial effectiveness of a television program. Efficiency, or cost per thousand home exposures, is selected as the criterion. The technique of sequential analysis is adapted to the telephone survey method to permit, with stated risks of error, a decision that the efficiency of a television program is better than a designated value, or worse than another designated value, or probably lies somewhere between the two limits. The combination of small cost and speed with which such efficiency decisions may be made suggests the use of a sequential decision plan by advertisers in many markets on a weekly basis as a tool in evaluating the effectiveness of their television advertising.

REFERENCES

Myers, L. An examination of television audience measurement methods and an application of sequential analysis to the telephone interview method. Doctoral dissertation, Syracuse Univer., 1956 (University Microfilms Publ. No. 56-3117).

Statistical Research Group, Columbia Univer. *Sequential analysis of statistical data: Applications.* New York: Columbia Univer., 1945.

Wald, A. *Sequential analysis.* New York: Wiley, 1947.

Wald, A. Sequential method of sampling between two courses of action. *J. Amer. Statist. Ass.*, 1945, **40**, 277–306.

(Received April 16, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

A NEW METHOD FOR CLUSTERING JOBS[1]

DAVID B. ORR[2]

Teachers College, Columbia University

During recent years, the United States Employment Service has been actively working toward the development of a truly functional job classification system. To this end, their researchers have analyzed 4,000 jobs, selected from the *Dictionary of Occupational Titles*, or *DOT* (1949), with respect to eight components or possible bases of classification.

Problem

The present study employed the data gathered by the United States Employment Service for these 4,000 jobs for only one component, aptitudes necessary for successful job performance. It employed the Distance (D) Measure, recently discussed by Cronbach and Gleser (1953), as a means of classifying jobs into clusters, relatively homogeneous with respect to the aptitudes measured. Two general questions were investigated: (a) the extent to which the D Measure offers a consistent, workable technique for clustering jobs on the basis of aptitude data; and (b) the extent to which the job clusters resulting from this clustering technique are consistent, reasonable, and describable in terms of outstanding aptitudes or patterns of aptitudes.

Thus, the present study was primarily exploratory. It dealt with only one of the components defined by the United States Employment Service, aptitudes necessary for successful job performance; it did not contemplate a complete approach to the more general problem of functional job classification.

The Statistic

The D Measure was chosen for study as a possible and relatively little explored approach to the general problem of classification and to the specific problem of job classification. Under the assumptions of: (a) equal unit scales, (b) equivalence of metric from one aptitude to another, and (c) independence among the aptitudes; the D Measure can be thought of as a generalization of the geometric formula for the distance between two points in a plane, viz.,

$$D_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

In the present study, these assumptions appeared to be reasonable.

The present study dealt with several samples of jobs, each of which had been rated as to the amount of each of nine aptitudes necessary for successful job performance. (The samples, rating scale, and aptitudes are described in a later section of this paper.) In the use of the D Measure with these data, each of the nine aptitudes was taken as an orthogonal dimension in higher dimensional space, and the distance concept was generalized to this higher dimensional case as follows: Let D_{ij} be the distance between Observations (jobs) i and j, and let a refer to any one of the m (aptitude) dimensions ($a = 1, 2, \ldots, 9$); let X_{ia} and X_{ja} be the score values (ratings) for Observations i and j, respectively, on Dimension a; then

$$D_{ij} = \sqrt{\sum_{a=1}^{m} (X_{ia} - X_{ja})^2}$$

The basic use of this statistic in the present problem was literally to say that jobs were or were not distant from one another in order that they might be divided into relatively homogeneous groups which would be relatively distinct from one another with respect to the underlying system of aptitude dimensions. This was done by considering jobs which exhibited small Ds among themselves to be similar and to be dissimilar to other jobs with which their Ds were large.

[1] The study upon which this paper is based was submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree at Teachers College, Columbia University. The author wishes to express his gratitude to Robert L. Thorndike, under whose guidance the study was done, for his help and encouragement.

[2] Now at the American Institute for Research, Washington, D. C.

TABLE 1

THE APTITUDES AND THEIR DEFINITIONS

Code	Aptitude
G	*Intelligence:* General learning ability: the ability to "catch on" or understand instructions and underlying principles; ability to reason and make judgments.
V	*Verbal:* Ability to understand meanings of words and ideas associated with them, and to use them effectively; to comprehend language and understand relationships between words and to understand meanings of whole sentences and paragraphs.
N	*Numerical:* Ability to perform arithmetic operations quickly and accurately.
S	*Spatial:* Ability to comprehend forms in space and understand relationships of plane and solid objects; may be used in such tasks as blueprint reading and in solving geometry problems; frequently described as the ability to "visualize" objects of two or three dimensions.
P	*Form Perception:* Ability to perceive pertinent detail in objects or in pictorial or graphic material; to make visual comparisons and discriminations and see slight differences in shapes and lengths.
Q	*Clerical Perception:* Ability to perceive pertinent detail in verbal or tabular material; to observe differences in copy, to proofread words and numbers, and to avoid perceptual errors in arithmetic computation.
K	*Motor Coordination:* Ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed; ability to make a movement response accurately and quickly.
F	*Finger Dexterity:* Ability to move the fingers, and manipulate small objects with the fingers, rapidly or accurately.
M	*Manual Dexterity:* Ability to move the hands easily and skillfully; ability to work with the hands in placing and turning motions.

PROCEDURE

The general procedure for the study was to take the aptitude data compiled by the United States Employment Service [3] and apply to it a clustering procedure, using the *D* Measure as suggested by Thorndike (1953).

Description of the Data. For the purposes of this study, aptitudes were defined as the specific capacities required of an individual in order to facilitate the learning of some task or job duty. Nine aptitudes derived from the United States Employment Service General Aptitude Test Battery, or GATB (1946–47), were used as the basic dimensions of this study. These nine aptitudes and their definitions are presented in Table 1.

The ratings of each job for each of the nine aptitudes indicated how much of each aptitude each job required for *satisfactory,* not maximum, performance. These ratings were made, according to the scale defined in Table 2, by six specially trained raters, using detailed job descriptions as their raw data. Each of these ratings was reviewed by other raters, and on the whole, no significant disagreement was found. The few cases of disagreement which did appear were resolved by the joint action of the rater and the reviewer.

Tratner, Fine, and Kubis (1955) have explored the intraclass reliabilities of the ratings for the nine aptitudes employed in the present study. These reliabilities ranged from 0.79 to 0.96 with a median value of 0.91. Validities of the ratings were on the order of 0.5 to 0.6 as measured against average GATB scores for successful workers in a number of occupations.

TABLE 2

THE DEFINITION OF THE RATING SCALE

1—An amount possessed only by the top 10% of the working population.
2—An amount possessed by the highest third, exclusive of the top 10%.
3—An amount possessed by the middle third of the working population.
4—An amount possessed by the lowest third, exclusive of the lowest 10%.
5—An amount possessed by the lowest 10% of the working population.

[3] The investigator would like to express his appreciation to Sidney A. Fine, Chief, Entry Occupations Section, Bureau of Employment Security, United States Department of Labor, and to the United States Employment Service for making these data available for the study.

TABLE 3

CLUSTER CENTROIDS AND SAMPLE JOBS FOR SIX CLUSTERS FOR SAMPLE A

Aptitude	Cluster Centroids					
	I ($N = 14$)	II ($N = 38$)	III ($N = 35$)	IV ($N = 10$)	V ($N = 20$)	VI ($N = 23$)
G	2.2	3.7	3.0	2.1	2.8	3.3
V	2.5	4.0	3.4	2.3	3.0	4.0
N	2.6	4.7	3.3	2.4	3.2	4.6
S	2.6	4.0	3.1	4.4	3.8	3.7
P	2.1	3.9	2.9	3.6	3.2	3.2
Q	3.9	4.9	4.0	2.4	3.6	4.7
K	2.6	4.0	3.4	3.3	4.0	3.4
M	2.6	4.0	3.1	3.5	4.0	3.6
F	2.9	3.7	2.9	3.7	3.8	3.0

Typical Jobs by Cluster

I: Meter Engineer, School Nurse, X-Ray Technician.
II: Cement Finisher Helper, Cooky Mixer Helper, Hydraulic Riveter Helper.
III: Tubular Furniture Maker, Transformer Tester, Steam Fitter.
IV: Rate Analysis Clerk, Public Accountant, Composer.
V: Coal Inspector, Service Establishment Attendant, Fish Inspector.
VI: Machine Driller, Stain Sander, Steam Table Attendant.

The Samples. The original, master sample of 4,000 jobs was chosen by the personnel of the United States Employment Service with regard for the representativeness of the sample on each items as code area in the *DOT*, variety of function, industry, and level of complexity. The final selection of the 4,000 jobs in the Master Sample was made according to the percentages of the sample thought necessary to represent each of the major code areas of the *DOT* adequately.

Two samples were used in the clustering procedure. These were completely independent, stratified, random subsamples of the Master Sample, drawn according to the same schedule of percentages as that used to select the Master Sample. These samples, hereafter called Samples A and B, were drawn with replacement and were of 140 jobs apiece. While no jobs appeared more than once in either sample, four members of Sample A were also members of Sample B.

Subsequently Sample C, consisting of 28 jobs, four chosen at random from each of the seven major code areas of the *DOT*, was drawn without replacement.

Treatment of the Data. The first step in the procedure was to determine an optimum number of clusters for the data. Then, each of the independent samples, A and B, was clustered independently for this optimum number of clusters. A statistic of consistency was devised to indicate the similarity of the two independent cluster structures. Finally, the clusters were compared for consistency and reasonableness of content.

The graphic procedure suggested by Thorndike (1953) was employed to determine the optimum number of clusters. No definitive decision could be made on the basis of this procedure, but six was the number of clusters tentatively identified as optimum for these data. Therefore, the remainder of the study dealt with the clustering of Samples A and B for six clusters.

The Clustering Process. The specific mechanisms of the clustering process followed Thorndike (1953) closely. This procedure results in clusters which are maximally compact over the entire structure, i.e., the mean distance for all observations from the members of their respective clusters is a minimum over all of the clusters. Then the mean score (rating) of the members of a cluster on each dimension is computed for each of the clusters. Where there are m dimensions, these m means taken collectively for a cluster are termed the *centroid* of the cluster.

The Statistic of Consistency. In order to determine the consistency of the two independent clusterings of Sample A and Sample B, it was necessary to devise a statistic. It was finally decided to match the six clusters from Sample A with the six clusters from Sample B, subject to the restriction that each cluster must have one and only one mate in the other clustering. This matching was accomplished by representing each cluster by its centroid and comparing the intercentroid distances using the D meas-

ure. Thus the mate for each cluster was the cluster in the other sample whose centroid was closest to its own.

Once the clusters were matched, each of the 28 jobs in Sample C was assigned to a cluster in both the Sample A clustering and the Sample B clustering. These assignments were made by computing the *D*s from these jobs to each cluster centroid and placing each job into the cluster to which it was closest. The statistic of consistency was the percentage of Sample C assigned in this manner to matched clusters.

Results

Presentation of the complete clusterings is too lengthy to include here, but Tables 3 and 4 present the cluster centroids (mean ratings for each aptitude) and several "typical" jobs from each cluster for the optimum ($k = 6$) clusterings for Samples A and B, respectively. Table 5 shows the extent to which the 28 jobs in Sample C, and also the four Sample A jobs which were repeated in Sample B, fell into matched clusters. Twenty-four of these 32 jobs or 75% fell into matched clusters, and 90.6% of them fell into the closest or next closest to the matched cluster in *both* samples.

The First Question

The first question underlying this investigation concerned the consistency and workability of the clustering technique based upon the use of the *D* Measure. Several results appear to bear on this question. The technique is consistent and workable to some extent because it was possible to match the clusters from Sample A to those derived independently from Sample B.

It is felt that, within the limitations imposed by the design and the assumptions underlying the technique, the percentage of jobs falling into matched clusters was good, and supports the workability and consistency of the technique. It is possible that those jobs falling outside of matched clusters might indicate another cluster. However, an examination of the jobs involved, and their ratings, does not seem to support this conjecture. It is more likely that these are jobs which fell at the boundaries of the present clusters and were moved into unmatched clusters through the action of the various kinds of error involved in the procedures.

Further evidence of the consistency of the

TABLE 4

Cluster Centroids and Sample Jobs for Six Clusters for Sample B

Aptitude	Cluster Centroids					
	I ($N = 12$)	II ($N = 31$)	III ($N = 10$)	IV ($N = 35$)	V ($N = 19$)	VI ($N = 33$)
G	2.0	3.8	1.5	2.9	2.9	3.2
V	2.7	3.9	1.7	3.8	2.9	4.0
N	2.2	4.8	3.0	3.1	2.8	4.2
S	2.1	4.0	4.2	2.9	4.0	3.5
P	2.2	3.9	3.6	2.7	3.7	3.2
Q	4.0	4.8	3.5	4.0	3.2	4.4
K	3.6	4.0	3.7	3.7	3.6	3.5
M	3.4	3.9	3.6	3.1	3.7	3.4
F	3.5	3.7	3.7	2.7	3.6	3.2

Typical Jobs by Cluster

I: Engineering Specification Writer, Pattern and Foundry Inspector, Wood Inspector.
II: Cameraman Assistant, Cooky Mixer Helper, Pipelayer Helper.
III: Actor, Law Examiner, Parole Officer.
IV: Shop Ironworker, Rock Cutter, Atomic Welder.
V: Beef Weighing Clerk, Companion, Depot Master.
VI: Cable Splicer, Hand Sewer, Thermite Welder.

TABLE 5

PLACEMENT OF SAMPLE C JOBS IN SAMPLE A AND B CLUSTERS

Job Number	Job Title	Clusters Sample A	Clusters Sample B
1	Bootblack	II	II
2	Cheese Maker	V[b]	VI[b]
3	Chitterling Splitter	II	II
4	Coremaker, Machine II	III	IV
5	Diesel Engine Erector	III	IV
6	Drafting Clerk	I[a]	IV[a]
7	Farm Mechanic	III	IV
8	Fisherman, Oyster	II	II
9	Fisherman, Shrimp	II	II
10	Geophysicist	I	I
11	Graphotype Operator	III[b]	V[b]
12	Establishment Guide	V	V
13	Hemstitching Machine Operator	VI	VI
14	Ingredient Scaler	V	V
15	Investigator I	V[a]	III[a]
16	Joggle Press Operator	III	IV
17	Lobby Man	II	II
18	Mgr., Trade-Mark, Copyright	I	I
19	Marker, Shipments I	V	V
20	Painter, Yachts, Small Boats	III	IV
21	Photographer, Commercial	I	I
22	Scalder	II	II
23	Second-Ride Fare Collector	III	IV
24	Sheep Farmer	III[a]	VI[a]
25	Size-Maker Helper	II	II
26	Soda Dispenser	III[a]	VI[a]
27	Solder Maker	III	IV
28	Zoologist	I	I
	Sample A Jobs Repeated in Sample B		
	Molder, Finish	III	IV
	X-Ray Technician I	I[a]	IV[a]
	Cooky Mixer Helper	II	II
	Reclamation Division Inspector	V[a]	IV[a]

Note.—Unfootnoted entries indicate that the job fell into matched clusters.

[a] Denotes jobs falling into the next closest clusters in *both* samples.

[b] Denotes any larger disagreements.

results is the fact that the results seemed to be unique. During the process of determining the optimum number of clusters, the cluster structure for Sample A for $k = 3$ was obtained twice, using vastly different initial clusters. Nevertheless, the redistribution steps of the procedure produced final cluster structures which were identical.

The Second Question

The second question concerned the reasonableness and describability of the clusters in terms of outstanding aptitudes or patterns of aptitudes. Evidence on this question may be had from an examination of the centroids presented in Tables 3 and 4, and from the face validity of the content of the clusters. Roman numerals identify the clusters in the order in which they emerged; primes signify Sample B clusters. Matched clusters are combined for descriptive purposes.

Cluster I-I'. This cluster is characterized by high level technical, supervisory, and mechanical jobs. The aptitudes required are Intelligence, Verbal, Numerical, Spatial, and Form Perception. Clerical is rated low.

Cluster II-II'. This cluster is composed of extremely low level jobs, primarily unskilled.

Cluster III-IV'. This cluster is composed largely of fairly high level skilled jobs, mostly mechanical and artisan types. The emphasis is on Spatial, Form Perception, Manual Dexterity, and Finger Dexterity.

Cluster IV-III'. The jobs in this cluster are of very high level with respect to their requirements of Intelligence and Verbal. Supervisory jobs are contained in this cluster, and there is some emphasis on clerical abilities.

Cluster V-V'. The emphasis here is on clerical jobs and supervisory jobs at a lower level than those in Cluster I-I'. The aptitudes required in the greatest amounts are Intelligence, Verbal, Numerical, and Clerical.

Cluster VI-VI'. This cluster features mechanical-manual jobs of a medium grade of skill. Intelligence, Finger Dexterity, and Form Perception are required, but no aptitude is needed in more than an average amount.

Although patterns of distinguishing aptitudes were not as clear as might have been desired, the content of the clusters seemed to support the hypothesis that the clusters were reasonable and describable in terms of patterns of aptitudes. An important finding is that the chief differentiations among clusters appeared to be both in terms of kind and level. Three different kinds of clusters appeared: intellectual-supervisory, mechanical-manual, and clerical. Within these types, several clusters were differentiated from others

chiefly on the basis of levels of aptitudes required.

Discussion

In general, the approach explored appears to hold some promise for dealing with the classification and functional structure of various job domains. In assessing the practical value of the approach, however, several points should be made.

The procedure as here employed is highly impractical for any sizable amount of data without recourse to mechanical data processing equipment for the computation of the *D* matrix, and for the computation of the mean intercluster and intracluster distances in the reassignment procedure.

Two limitations of the *D* Measure itself should be mentioned. First, being based upon a sum of squared differences, the *D* Measure takes no account whatsoever of the direction of those differences. Second, no *D* value is unique; it could be made up of a large difference between the two jobs on only one dimension, or of smaller differences on all of the dimensions, or of any state of affairs in between. These factors suggest that the *D* Measure may misrepresent the facts, to some extent, blurring the distinction between kind of job and level of job.

In addition, certain other factors limit the interpretation of the results. First, the samples used were small, and those small groups of jobs requiring specific or high level aptitudes tended not to be found in the same proportions, if at all, in both samples. Second, the amount of spread in the ratings was small, thus almost certainly curtailing the discriminatory power of the technique. Finally, it must be remembered that the cluster structures were the result of an averaging process, and thus some jobs could fit hardly at all in their clusters, and yet fit there better than elsewhere.

Summary

Aptitude ratings made by the United States Employment Service of the amount of each of nine aptitudes necessary for success in certain jobs were employed to explore the effectiveness of the Distance (*D*) Measure as a statistic for the clustering of jobs. Two questions were considered: (*a*) the extent to which the *D* Measure offers a consistent, workable technique for clustering jobs on the basis of aptitude data; and (*b*) the extent to which the job clusters resulting from such a clustering technique are consistent, reasonable, and describable in terms of outstanding aptitudes or patterns of aptitudes.

Six was found to be the best number of clusters for these data, and two independent samples of 140 jobs apiece were each clustered independently for six clusters. Tables showing the cluster matching and cluster centroids were presented, and a statistic of consistency between the two samples was developed and applied. The content of the clusters exhibited show a differentiation in terms of intellectual-supervisory, mechanical-manual, and clerical jobs, with further differentiation on the basis of level of aptitudes required for success.

It was tentatively concluded that the use of the Distance Measure, and the clustering technique employing it to define job structures showed sufficient promise to warrant further investigation.

REFERENCES

Cronbach, L. J., & Gleser, Goldine C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.

Thorndike, R. L. Who belongs in the family? *Psychometrika*, 1953, **18**, 267–276.

Thorndike, R. L., Hagen, Elizabeth P., Orr, D. B., & Rosner, B. An empirical approach to the determination of Air Force job families. USAF Pers. Train. Res. Cent., *Tech. Rep. AFPTRC-TR-57-5*. August 1957, Lackland Air Force Base, Texas.

Trattner, M. H., Fine, S. A., & Kubis, J. F. A comparison of worker requirement ratings made by reading job descriptions and by direct job observation. *Personnel Psychol.*, 1955, **8**, 183–194.

United States Employment Service. *General Aptitude Test Battery*. Washington, D. C.: United States Government Printing Office, 1946–47.

United States Employment Service. *Dictionary of Occupational Titles*. Washington, D. C.: United States Government Printing Office, 1949.

(Received April 30, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

THE EFFECTS OF SLEEP LOSS ON COMPONENT MOVEMENTS OF HUMAN MOTION[1]

SHELBY J. HARRIS

Lehigh University

The major purpose of this experiment was to evaluate the effects on component movements in skilled motions of stress produced by prolonged sleep loss. Studies in which the effects of physiological stresses on human motor performance have been investigated have usually employed a variety of tests which yield measures of overall accomplishment, such as number of units completed or number of errors per trial. In the present experiment the major emphasis is on the evaluation of changes occurring in different types of movements within the same motion pattern. Specifically, the study was primarily concerned with the effects of 60 hours of sleeplessness on the duration of travel and manipulation movements in patterned motions.

The value of component analyses of motion in relation to stress was suggested by an earlier investigation of prolonged food restriction. In this study (Harris, Brozek, & Smith, 1958) it was shown that a 24-day period of maintenance on a 1,000-calorie, all-carbohydrate diet produced a deterioration in gross movements of the arms and body (travel). No systematic change in the finer manipulative movements of the fingers, wrist, and lower arm was evident. It is of interest to determine whether the differential effect on component movements demonstrated in the food restriction study is specific to nutritional deficit or is a more general reaction to psychophysiological stresses of various types.

In addition to a panel-control task used in appraising the duration of motion components, the present study included four other tests of perceptual and motor functions.

[1] This paper is based on a thesis submitted to the Department of Psychology of the University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The research was supported by funds granted by the Graduate School Research Committee, University of Wisconsin.

METHOD

The general plan of this experiment consisted of the administration of several special tests of perceptual and motor functions in a 5-day training period, in a 3-day sleep loss period, and in a 2-day recovery period.

Prior to the sleep deprivation period each *S* was given five consecutive days of practice on all of the tests used in the study. Data were recorded only on the last two days of the training period. In order to evaluate diurnal variations in performance during the three-day period of sleeplessness, tests were administered in the morning (5–8 A.M.), about noon (11 A.M. to 1 P.M.), in the early evening (5–8 P.M.), and midnight (11 P.M. to 1 A.M.). All tests, with the exception of the panel-control task, were used in each test session. To lessen the risk of *S*s developing finger blisters, the panel task was included only in the morning and evening sessions.

All of the tests were administered again once on each of the two days immediately following the sleep loss period. With two exceptions the test sessions took place in the afternoon, usually between 3 and 6 P.M. In two cases the tests were administered in the latter part of the morning on the first recovery day.

The *S*s were closely supervised during both nights and the last two days of the sleep loss period, except at mealtimes when they were allowed to return to their living quarters for an hour or less without supervision. On the second day of the sleep loss period, some *S*s were also permitted to leave the laboratory in order to attend classes, perform meal jobs, carry out necessary activities at other labs and, on one occasion, to attend a football game. The *S*s typically spent their time in rather light activities such as studying, playing cards, throwing darts, taking short walks, etc.

The *S*s were allowed to sleep as long as they desired on the nights preceding and following the sleepless period. On the night immediately preceding sleep loss, *S*s slept an average of about 7.5 hr. Immediately following sleep loss an average of 12.3 hr. of sleep was taken.

No independent control group was used in the study. It was decided to use the data from the first sleep loss day and the recovery period as a baseline to evaluate changes accompanying sleep loss. This decision to use *S*s as their own controls was based on several considerations. Previous data on the panel-control test, which was the task of primary interest in the study, indicated that the amount of training given during the training period should be adequate

to lead to asymptotic performance on both motion components (Harris et al., 1958). Further, the practical difficulties in obtaining voluntary *Ss* to submit to the rigors of the experiment and the unfeasibility of testing *Ss* at 6-hr. intervals without interfering with their sleep schedules dictated against the use of an independent control group.

Subjects. Nineteen male students at the University of Wisconsin served as *Ss*. Eight *Ss* were graduate students and 11 were undergraduates enrolled in psychology courses. The *Ss* ranged in age from 19 to 26 yr. The *Ss* participated in groups of three and, on one occasion, four individuals.

Tests and Test Procedures

Panel-Control Test. The apparatus and testing procedure used for assessing the duration of travel and manipulation movements in a motion pattern has been previously described (Smith & Wehrkamp, 1951). Briefly, *S*'s task consisted of turning a series of 28 rotary switches mounted on a control panel. Separate measurement was obtained of the time involved in turning (manipulation) and in moving between switches (travel). Two conditions of performance were investigated: a 10-min. period of continuous performance at a "normal" or "comfortable" pace and a 2-min. period of continuous performance at maximum pace. These conditions were the same as those employed in the food restriction study (Harris et al., 1958).

Bimanual Coordination. This test consisted of four columns of five two-position, toggle-type switches. In bimanual performance *S* worked back and forth with the right hand between the two columns of switches on the right and, at the same time, with the left hand between the two columns of switches on the left. The distance between the outside and inside columns was 18.5 in. Total time involved in flipping the 20 switches was automatically registered on an electric timer.

Two bimanual patterns of travel motion were used. These were "simultaneous" and "mirrored" motions. In the simultaneous motion patterns, the two arms moved between the columns of switches in the same direction at the same time. In mirrored motions, the two arms moved simultaneously in opposite directions between the columns of switches. Time-scores for five subtests of each type of motion pattern were summed and treated as a single score.

Two tests of unimanual performance were also included. The task was identical except that the two hands were used singly. One test involved performance with the right hand and the other, performance with the left hand.

Instructions were given to work as rapidly as possible without making any errors. Repetition of a subtest was required if an error was made.

Leg Movement. The apparatus used in measuring the speed of a discrete leg movement was patterned after that developed by Brozek, Simonson, and Keys (1952). The *S*'s task was to throw his right leg straight up as far and as fast as possible. The duration of the movement from the instant the foot leaves the floor to the point where the thigh is horizontal was recorded. Three trials were given at each test session.

Steadiness. In this test *S*'s task was to hold a rod, ¼ in. in diameter, as steady as possible with his right hand while resting his elbow on a table. Lateral deviations of the rod illuminated one of two lights indicating the direction of *S*'s deviation. The number and duration of deviations from the center position were recorded. Three 30-sec. trials were given at each test session.

Critical Flicker Frequency. A circular test patch ⅜ in. in diameter was viewed at a distance of 7 in. The brightness level of uninterrupted light beam was maintained constant at about 5.5 mL. The light-dark ratio was one, and monocular vision was used throughout. Unlimited time was allowed for *S* to make the adjustment to the point of fusion. Eight trials were given at each test session with alternate trials starting at flicker rates of varying amounts above and below fusion.

Order of Tests. The following order of administration of the tests was used in each session: (*a*) steadiness, (*b*) CFF, (*c*) bimanual coordination, (*d*) leg movement, and (*e*) panel-control test.

Results

Figure 1 shows the mean performance level for the manipulation and travel movements under the normal-pace condition. The movement durations are shown for each of the test sessions during sleep loss (SL), the last two training days (T) and the two recovery days (R). Thus, the figure pictures the general level of performance for each of the 10-min. work periods without consideration of changes in performance within work periods. Duncan's

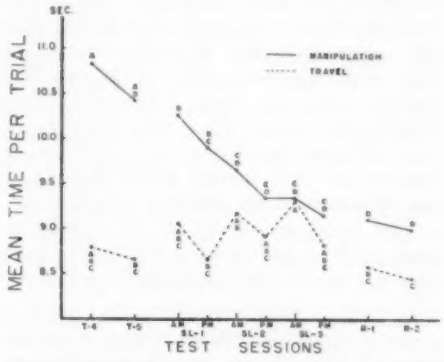

Fig. 1. Mean time per trial for the manipulation and travel movements under normal pacing at each of the test sessions during training (T), sleep loss (SL) and recovery (R) period. Any means which do not have a letter in common differ significantly.

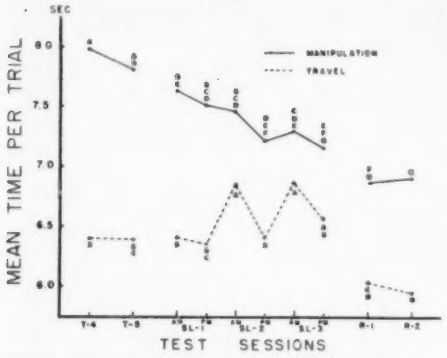

FIG. 2. Mean time per trial for manipulation and travel movements under maximum pacing at each of the test sessions. Means which do not have a letter in common differ significantly.

multiple-comparisons test (1951) was applied to the manipulation and to the travel time data.[2] The results of the Duncan test are also shown in Fig. 1. Letters are used to indicate groups of means which do not differ significantly. Any pair or group of means which do not have a letter in common are significantly different.

Examination of the results of the Duncan test shows that the decrease in manipulation times during sleep loss is significant. The manipulation values do not return to the level of control values on the two recovery days.

The results of the Duncan test on the means of the travel movements demonstrate that the increase in travel times during sleep loss is not significant. Only the morning session on SL-3 exceeds either of the two SL-1 control values and, there, it exceeds only the evening test session. Thus, only when the effects of sleep deprivation are considered in conjunction with diurnal effects do any of the differences attain statistical significance.

The analogous curves for the maximum-pace condition are shown in Fig. 2. The mean time per trial for the manipulation and travel movements during the 2-min. work period is plotted for each of the test sessions. In contrast to the results on normal pace, the pronounced diurnal effects on travel movements under maximum pacing become apparent only after the loss of one night of sleep. Further, the mean travel times in the evening sessions continue to increase throughout the three days of sleeplessness. The travel times on the two recovery days are below the values on the first sleep loss day.

The results of the Duncan test indicate that the general decline in manipulation times over the sleep loss and recovery periods is significant. The duration of the travel movements are significantly greater in the morning test sessions of the second and third days of sleep loss. While the travel times in the evening of the second and third sleepless days are not significantly greater than on SL-1, they are significantly greater than on either of the recovery days.

In order to evaluate changes in movement times within work periods, as contrasted to changes between work periods, the medians of the first three, middle three, and last three trials of each work period were selected for analysis. These medians were computed, separately, for the two motion components on the two conditions of pacing. Only the data from the test sessions during the sleep loss and recovery periods were included in the analyses.

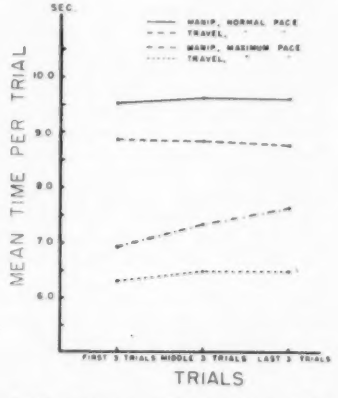

FIG. 3. Mean time per trial for the medians of the first three, middle three, and last three trials of each work period. Means shown are for the pooled data of the sleep loss and recovery periods.

[2] In all the analysis of variance and Duncan tests, significance at the .01 level was required before significance at the .05 level was accepted. This more conservative requirement seemed warranted because of the possible departure from normality of some of the data. Summaries of analyses of variance may be found in the thesis (Harris, 1955) upon which this paper is based.

In no instance was the Trials × Test session interaction significant. This lack of significance indicates that changes in component movement times within work periods (i.e., work decrements) were not consistently altered during the course of the sleep loss and recovery periods.

Figure 3 shows the changes in component movement times within a work period under the two conditions of pacing. The data for the sleep loss and recovery periods have been pooled to obtain the means shown in the figure. Under normal pacing, neither of the motion components shows a significant change over trials. On the other hand, both manipulation and travel times increase significantly during the 2-min. work periods at maximum pace.

Bimanual Coordination. Figure 4 shows the mean total time for 16 *Ss*[3] on all 10 of the subtests of bimanual performance. The data for the mirrored and simultaneous motion patterns have been pooled to obtain the means shown in the figure. This pooling of the data is justified by the lack of significant differential effects of sleep loss on the two types of motion patterns. Mirrored motions were superior to simultaneous motions throughout the experiment.

Comparison of the mean performances in morning and evening sessions shows that the diurnal changes in performance are significant. With respect to the changes in performance attributable to sleep loss, only the morning session of SL-3 shows a significant deterioration when SL-1 is used as the basis of comparison. However, performance on both recovery days is significantly superior to that of all test sessions of SL-3.

The mean total times to perform the two unimanual tests are also shown in Fig. 4. The rationale for pooling the data from the two hands is the same as that for combining the data of the two motion patterns in bimanual performance, i.e., there is no significant differential effect of sleeplessness on performance by the two hands. As may be seen in Fig. 4, the changes in unimanual performance

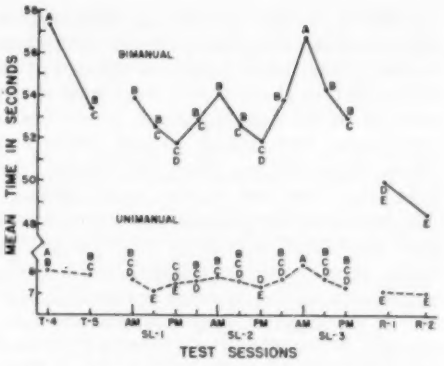

FIG. 4. Mean total time for the 10 subtests of bimanual and for the two subtests of unimanual performance. Means which do not have a letter in common differ significantly.

are very similar to those of bimanual performance.

Leg Movement.[4] The median duration of the three leg movements performed by each S at each test session were used in the analysis of the data. The mean of these median scores was then computed. The results of the Duncan test indicated that all three means on SL-3 were significantly above the SL-1 values. These same means were not significantly higher than the mean on R-1 but were reliably greater than on R-2. The decrease in speed of a simple leg movement occurred entirely during the second night of sleep loss.

Steadiness. The results on the test of hand steadiness were evaluated both in terms of number of "contacts" (deviations from the center position) and mean duration per contact. No reliable change in mean duration of contacts over test sessions of the experiment was demonstrated in analysis of variance of this data.

In the analysis of the number of contacts, test sessions proved to be a significant source of variation. However, the change in number of contacts as a function of sleep loss was irregular. Only the means of the midnight session on SL-2 were significantly below the control values.

[3] Due to a change in the procedure in administering and scoring the test of bimanual and unimanual coordination, the first group of three *Ss* was not included in the analysis of the data.

[4] For the sake of brevity, only a brief description of the results on the leg movement, steadiness, and CFF tasks is presented here. A more detailed account is available elsewhere (Harris, 1955).

Critical Flicker Frequency. The median flicker fusion frequencies of the last three trials starting below fusion and the last three trials starting above fusion were used in the analysis of the data. Inasmuch as the fusion frequencies from initial settings above and below fusion did not differ significantly, the data from these two sources were combined. The results of the Duncan test demonstrated that the overall decline in CFF during the 60-hr. sleep loss was significant. However, the significance of the decrease produced by the loss of one night of sleep was somewhat ambiguous. While all of the values of SL-2 were significantly below all means on SL-1, they were not significantly below the value on R-1.

Discussion

The absence of an independent control group and the failure of the performances during the recovery period to return to the presleep loss levels make the interpretation of some of the data somewhat equivocal. In the CFF, steadiness and leg movement tasks and in the travel movements under normal pacing, the level of performance is essentially the same on SL-1 and during recovery. However, the bimanual coordination test, both motion components on the panel-control test under maximum pacing and the manipulation movements under normal pacing show a superior performance during the recovery period.

Within the design of the present experiment, the possibility that on some tasks detrimental effects of sleep loss have been obscured or diminished by continued learning during the sleepless period cannot be precluded. In the case of the bimanual coordination test, where no previous learning data is available, there is no objective evidence against this possibility. However, on the panel-control task, if the results are considered in relation to earlier studies, it seems unlikely that practice alone is adequate to account for the improved performance on R-1 and R-2. Previous studies (Harris et al., 1958; Smith & Smith, 1955) indicate that the amount of training given prior to the sleepless period should have been enough to lead to asymptotic or near-asymptotic performance on both motion components.

It seems more likely that the improved performance is largely attributable to compensatory motivation accompanying sleeplessness. While no systematic ratings of "effort expended" were obtained, general observations of *S*'s test behavior and comments strongly suggested that the tests became more challenging as sleeplessness progressed. If the higher standards of performance established late in the sleepless period carried over to the recovery period, this would account for the superior performance after sleep loss. It should be noted that improvement in performance during and after sleep loss occurs, primarily, in those tests in which self-pacing is an important aspect of the task. This would be expected if compensatory motivation were an important factor.

A differential effect on manipulation and travel movements produced by a psychophysiological stress seems to be clearly demonstrated in the panel-control task. This result is in agreement with the earlier study in which the same task was used to appraise performance changes accompanying food restriction. In both instances the deleterious effects of the stress condition were confined to the travel movements.

The data on travel movements in the panel-control task and the unimanual and bimanual coordination tasks, in which travel movements predominate, show consistent diurnal variations during the three-day sleep loss period. These observations, in conjunction with those noted above, suggest that variations in "physical fitness" produced by psychophysiological variations are reflected primarily in repetitive travel movements rather than in other types of work movements. Such movements seem to be sensitive indicators of normal psychophysiological variations as well as of the more extreme stress conditions produced by sleep loss and nutritional deficit.

Summary

In this study the effects of 60 hr. of sleeplessness on various psychomotor and perceptual performances have been investigated. The major purpose of the experiment was to determine whether sleep loss produces differential effects on component movements of

motion. The results may be summarized as follows:

1. The loss of sleep had a differential effect on manipulation and travel movements in a panel-control task. The duration of manipulation movements tended to decrease during sleep loss. Travel movements, particularly under a maximum-pace condition, tended to increase in duration during the sleep loss period.

2. Speed of performance in a test of bimanual and unimanual coordination gave results similar to those obtained for travel movements in the panel-control task.

3. Speed of a discrete leg movement decreased during the sleepless period.

4. The number of contacts in a test of hand steadiness showed an irregular change during the sleep loss period. No significant change in the duration of contacts was obtained.

5. A significant decrease in CFF was found during sleeplessness.

6. The data on travel movements in the panel-control task and the unimanual and bimanual coordination tasks showed consistent diurnal variations during the three-day sleep loss period.

REFERENCES

BROZEK, J., SIMONSON, E., & KEYS, A. A test of speed of leg and arm movements. *J. appl. physiol.*, 1952, **4**, 753–760.

DUNCAN, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, **2**, 171–189.

HARRIS, S. J. The effects of sleep loss on component movements in human motion. Unpublished doctoral dissertation, Univer. of Wis., 1955.

HARRIS, S. J., BROZEK, J., & SMITH, K. U. The effects of caloric restriction on the travel and manipulation components of human motion. *Internat. A. angew. Physiol. einschl. Arbeitsphysiol.*, 1958, **17**, 34–41.

SMITH, PATRICIA A., & SMITH, K. U. Effects of sustained performance on human motion. *Percept. mot. Skills*, 1955, **5**, 23–29.

SMITH, K. U., & WEHRKAMP, R. A universal motion analyzer applied to psychomotor performance. *Science*, 1951, **113**, 242–244.

(Received May 11, 1959)

Journal of Applied Psychology
Vol. 44, No. 1, 1960

UNION AND MANAGEMENT TRAINEES—A COMPARATIVE STUDY OF PERSONALITY AND OCCUPATIONAL CHOICE[1]

HOWARD M. BOGARD[2]

Teachers College, Columbia University

Considerable research has been carried out in the area of investigating whether occupational personalities exist and whether certain kinds of people tend to enter specific occupations. Though studies have frequently failed to support a personality theory of occupational choice, such negative findings appear not completely to disprove the theory, but seem, in many instances, to have been the result of inappropriate instruments, loosely defined occupations, and failure to precede the selection of tests by adequate study of the occupation, with an emphasis on the personality demands of the job.

During recent years the literature has revealed an increasing concern by corporation executives with the problem of management succession and with appraisal of candidates for advancement. Concern with executive replacements has also been voiced by a group of labor leaders who have noted the trend toward professionalization in management and in their own unions.

In an attempt to study further the relationship of personality to occupational entry, the present investigation undertook to determine whether individuals considered by top union and management executives to be worthy material for training for leadership positions were different from each other in psychological areas.

[1] This study is based upon a dissertation submitted to Teachers College, Columbia University in partial fulfillment of the requirements for the PhD degree.

[2] The author would like to express his appreciation to the members of his committee: Donald E. Super, Henry David, Joel R. Davitz, and Albert S. Thompson. Acknowledgment is also made to the International Ladies Garment Workers Union and the Institute of Management and Labor Relations of Rutgers University for financial support. Gratitude is similarly due to the Grace Line, Inc., and again to the ILGWU for making their management trainees available for study.

Background and Procedure

In 1950, a formal leadership training institute was inaugurated by the International Ladies Garment Workers Union (ILGWU). Students accepted for the one year course of training and who are hopefully considered to be the future leaders of the union are, upon successful completion of their studies, guaranteed placement within the union in positions of responsibility, depending upon their specific skills and union needs at the time. The 12-month training course is divided between full-time class attendance and field work under supervision of incumbent officers.

Though carefully defined job specifications have not been constructed, the admissions committee has set forth general entrance requirements. At the present time candidates are generally expected to be, most of all, "dedicated to trade union ideologies." Those accepted are expected to have a "strong social feeling," to exhibit a willingness to remain relatively detached from close family ties, and to aver acceptance of the fact that their earnings will rarely, if ever, compensate them adequately for the hours that they will be required to work. Intelligence, level of formal education, and previous employment history are usually not considered too important by the union's admissions committee.

Job history, age, education, and military service are not verified, and the desired personality attributes are sought through a few rather brief though intensive screening interviews, and one appearance before the executive board of the union.

Nineteen male students in the ILGWU class graduated on May 30, 1957 and 21 students in the union class who completed their training in May of 1958 were selected to comprise the experimental group.

The management group chosen, in order to be considered comparable, had most of all to be comprised of trainees motivated to eventual leadership. As controls, it was desired that they be undergoing training for leadership, should be in an early stage of their job careers and should be as close as possible to the union group's age level. Intelligence and level of formal education, though of some relevance, were considered an important variable only if the groups investigated were radically different.

At its headquarters in New York City, the Grace Line, Inc., a major international shipping corporation, conducts a systematic and well administered management training program. They recruit widely for candidates and the reported acceptance ratio of applicants is low.

Though exceptions are made, superior intelligence,

high energy level, strong aggressiveness, and personal maturity are generally regarded to be the most desired characteristics in applicants. Age and marital status are not considered to be of great importance. Candidates are briefly screened by the personnel department, and those favorably considered are given a brief personality test. For those who successfully pass this hurdle, a full day of interviews with eight or nine officers of the corporation is then scheduled, and the decision of whether to accept or reject an applicant is made only after he has completed all of these steps.

Forty males employed by the Grace Line comprised the control group. Of these, sixteen were currently enrolled in the training program, eighteen were recent trainees and six were new executive employees.

Choice of Study Instruments

Segal (1953), in discussing techniques of studying occupations, emphasized the need to augment traditional job analysis methods by stressing the personality demands of the job. He suggested that the more relevant factors would then be exposed and made available for study.

With such an orientation in mind, and based in part on the "ideal types" sought by the union and management training program administrators and on material found in published biographies of union and management leaders, a "typical" trade union and management leader was drawn. Then, after proposing that specific kinds of early life experiences were significant in determining presently revealed personality traits, four testable hypotheses built upon the "typical" trade union and management leaders and on the noted approach to personality formation were formulated.

It was assumed that differences would be most apparent in the areas of:

1. Social class identification
2. Values
3. Personality characteristics
4. Biographical and developmental experiences

Due to *Ss*' time-consuming instructional program, individual interviewing was not possible and group administration of all tests was mandatory. After considering the required administration time and the vocabulary of published tests which seemed particularly well suited to investigating the above four hypotheses, the *Sims SCI Occupational Rating Scale*, the Allport-Vernon-Lindzey *Study of Values*, and the *California Psychological Inventory* were selected to study the first three areas.

As it was the intention of the study to investigate and describe in some detail "what kind of people" chose to enter careers in trade unions and management, an instrument which provided details of prior life experiences hypothesized to be related to specific personality trait development appeared useful.

An analysis, within a psychoanalytic framework, of the trade union and management "ideal types" enabled the postulation of certain early life experiences and family constellations which seemed basic to present behavior. Those were then refined into the following five variables: [3]

1. Relationships with authority figures
2. Maternal control of the family
3. Age and severity of independence training
4. Success in peer group relationships
5. Social conformity of the family

Items pertaining to these variables were constructed and then submitted to a group of judges for screening in order to increase the face validity. Employing the scores from both the union and management groups, odd-even reliability coefficients of the individual categories were computed and corrected by the Spearman-Brown formula, with all traits scoring sufficiently high to warrant confidence in

[3] Items pertaining to the first four variables were directed to surveying such areas as: (1) the kinds of experiences the respondents had with their parents, teachers, employers, etc.; (2) which of their parents exerted the most significant control over the family; (3) the age at which the respondent was first given certain responsibilities, was permitted to have dates, seek part-time employment, etc.; (4) the status held by the respondents in their play groups, clubs, teams, and formal organizations.

Social conformity, (5), was appraised by first tabulating various aspects of behavior, as recalled by *Ss*, of their immediate neighbors during childhood and adolescence and then later in the inventory, presenting the selected items in a different order and having *Ss* indicate what *their* family did in these situations. The discrepancy score, between the behavior of neighbors and the respondents' families, was employed as the indicator of Social conformity.

TABLE 1

EDUCATION OF UNION AND MANAGEMENT TRAINEES AND THEIR FATHERS

(N = 40 in each group)

Grade Completed	Union		Management	
	Trainees	Fathers	Trainees	Fathers
4 years college	8	0	38	13
3 years college	3	0	2	1
2 years college	8	3	0	3
1 year college	5	0	0	2
High school graduate	11	9	0	10
Less	5	28	0	11

the instrument for group use.[4] This custom-built *Biographical Inventory* comprised the fourth study instrument.

The average union age of 24.83, with a range of 20–40 years, when compared to the management mean of 27.75, with a range of 22–40 years, is so close as to suggest that any intergroup differences on the tests could not be assumed to be a function of an age discrepancy.

As the Grace Line places a premium on educational attainment, it is not surprising to note in Table 1 that the management group generally appears to have completed considerably more formal education than the union trainees. Though the administrators of the ILGWU training program suggest that amount of formal education is relatively unimportant, 60% of the union trainees state that they completed at least one year of college training. Almost half of the ILGWU students allege that they had completed two or more years of college and many suggest that financial need was the main reason for their not continuing their formal education. Furthermore, due to the professional approach of the institute staff and the strong academic content of their training, the possibility that some of the ILGWU students considered their attendance as a form of college or professional schooling should also be considered.

Two factors should be noted in considering

[4] Authority relationships .90
Material control .62
Independence training .58
Peer group relationships .83
Social conformity .76

TABLE 2

MARITAL STATUS OF UNION AND MANAGEMENT TRAINEES

Group	Single	Married	Total
Union	26	14	40
Management	18	22	40

the greater number of married Grace Line trainees (Table 2). A number of the management group had been married for only a very few years and as their mean age is almost three years greater than that of the ILGWU group, the psychological meaning of the difference in marital status may not be as striking as it appears. During the selection interviews, the ILGWU screening committee stresses a willingness of the candidates to remain relatively detached from close family ties. Those who were married and were fathers were apprised of the long and irregular hours that they would be called upon to work and particular emphasis was laid upon the fact that they could, upon completion of their training, be assigned to any area of the United States. Additionally, the applicants were provided with details of vigorous antilabor activities and tales of physical violence which had occurred in some of the communities to which they could be assigned.

Discussions pertaining to why certain people select union rather than management careers frequently raise the question of whether "their fathers were union members"? Table 3 indicates that 80% of the fathers of the ILGWU trainees were reported as having held membership in a trade union as opposed to only 32% of the fathers of the Grace Line trainees.

TABLE 3

UNION MEMBERSHIP OF FATHERS OF UNION AND MANAGEMENT TRAINEES

(N = 40 in each group)

Group	Union Member Yes	Union Member No
Union	33	7
Management	13	27

The educational achievement of the trainees' fathers (Table 1) significantly favors the management group. Though it is unlikely that there is a very high relationship in these particular groups between formal education and the intellectual or political atmosphere of the home, the fact that 47% of the fathers of the management group completed at least one year of college as opposed to only 7% of the union group's fathers is worthy of note.

As the present study was intended to look for intergroup differences, intellectual ability was considered as a variable to be evaluated in the same light as personality or developmental differences.

Prior to the administration of the regular test battery, all *Ss* included in the study were given the *Wonderlic Personnel Test*. Considering the marked intergroup differences in formal education and the preference given to more intelligent applicants by the Grace Line, the higher mean score (Table 4) of the Grace Line *Ss* is not surprising.

In considering the Wonderlic test scores as a measure of intelligence it should be noted that the union *Ss* were at somewhat of a disadvantage as the test, which is constructed along lines of increasing item difficulty, includes a number of questions toward the end which would be particularly easy for individuals who have had advanced training in mathematics. As the majority of the Grace Line trainees reported that they received their undergraduate training in engineering and business administration their observed superiority is perhaps not as significant as it initially appears.

The observed differences in intelligence test scores of the experimental and control groups raised the possibility that intergroup differences in personality, values or biographical data may be a function of intelligence rather than the union-management variable. Gough (1957), however, found almost no relationship between intelligence and scores on the *California Personality Inventory;* and correlation ratios computed between the *Wonderlic* scores and the *Biographical Inventory* categories and the *Study of Values* traits disclosed no significant ratios, with the exception of *social* values (−.42). Though this negative relationship between *social* values and intelligence is in disagreement with other published studies (Pintner, 1933), the correlation was taken into consideration in evaluating the total study results.

TABLE 4

WONDERLIC PERSONNEL TEST SCORES OF UNION AND MANAGEMENT TRAINEES

(N = 40 in each group)

Group	Mean	*SD*	Range
Union	24.9	5.56	14–35
Management	31.5	6.42	17–45

On the basis of the ILGWU job specifications and job duties and the general stereotypes of trade union and management leaders, 15 hypotheses were formulated:

As measured by the Sims SCI Occupational Rating Scale

1. The union group identifies with a lower *social class* than does the management group.

As measured by the Allport-Vernon Study of Values

2. The union group reveals higher *social values* than does the management group.
3. The union group demonstrates lower *economic values* than does the management group.

As measured by the California Psychological Inventory

4. The union group scores lower on *self-control* than does the management group.
5. The union group scores higher in *psychological-mindedness* than does the management group.
6. The union group scores higher in *responsibility* than does the management group.
7. The union group scores higher in *socialization* than does the management group.
8. Both the union and management groups score high in *dominance*.
9. Both the union and management groups score high in *self-acceptance*.
10. The union group scores higher in *femininity* than does the management group.

As measured by the Biographical Inventory

11. The union group has had a greater number of oppositional *relationships with authority figures* than does the management group.
12. The union group has had a greater number of rejecting experiences in *peer group relationships* and social participation than does the management group.
13. The union group, during childhood and adolescence, was supervised by a family which, in contrast to the management group,
    a. was *maternally controlled*.
    b. stressed *social nonconformity*.
    c. valued early *independence training*.

TABLE 5

MEANS AND STANDARD DEVIATIONS OF UNION AND MANAGEMENT GROUPS ON EACH OF THE HYPOTHESES TESTED

(N = 40 in each group)

Trait[a]	Union		Management		Student's t
	X	SD	X	SD	
1. Social class identification	16.97	5.67	25.84	4.07	8.045**
2. Social values	46.27	5.96	32.00	7.08	9.754**
3. Economic values	37.75	7.71	45.95	5.78	5.387**
4. Self control	27.65	7.52	29.25	7.03	.982
5. Psychological mindedness	12.73	2.63	13.33	2.15	1.110
6. Responsibility	29.78	3.99	32.65	4.77	−2.917[b]
7. Socialization	33.97	5.55	36.02	4.30	−1.846[b]
8. Dominance	33.20	4.38	34.90	1.64	[c]
9. Self acceptance	23.72	3.01	24.30	8.64	[c]
10. Femininity	16.75	3.01	15.50	2.99	1.862*
11. Oppositional relationships with authority figures	12.19	7.96	6.22	3.79	4.279**[d]
12. Rejecting peer group relationships	3.98	2.82	2.03	1.59	3.806**
13. Maternal control	6.13	3.47	5.87	3.23	..338[d]
14. Nonconformity	4.55	3.23	2.70	2.25	2.932**
15. Early independence training	8.77	2.63	6.20	2.84	4.208**

* Significant at less than the .05 level.
** Significant at less than the .01 level.
[a] The instruments employed for the measurement of each trait are indicated in Footnote 4.
[b] Since a one-tailed critical region gives no protection against alternatives in the direction opposite to the one hypothesized, it is obviously possible that such an alternative might be true and not be revealed by the presently employed one-tailed test. The computed t score, which is less than the first percentile, would suggest that the management group can be described as scoring higher than the union students in these traits.
[c] The hypotheses pertaining to Traits 8 and 9 did not specify intergroup differences, but hypothesized that the groups would score high when compared with other occupational groups. Comparison with published norms supported the hypotheses.
[d] N = 38.

On each of the tests included in the battery, the mean and standard deviation of the scores of each group were computed. The significance of observed differences was determined by t tests employing a one-tailed critical region. The maximum level of significance allowed was .05.

RESULTS

Generally the results (Table 5) justify the conclusion that the two groups are quite different. The union trainees are somewhat less intelligent and less aggressive than the Grace Line group, although both groups tend to score above the general population in both of these traits. They are more altruistic than the management trainees, less concerned with doing things that are practical, and reveal a greater number of interests usually found in individuals who deal with people, art, music, and languages as opposed to those who usually deal with objects and office records. Their social class identification is significantly lower than that of the management group. The ILGWU trainees, in light of the results, appear to be less responsible, tend to manifest less social maturity, and reveal a long-standing propensity to overt conflict with authority figures.

During their formative years, the management trainees appear to have been more successful in their peer group relations, to have received less early independence training and to have been reared by parents who were generally more conforming to social pressures than were those of the ILGWU students, whose fathers received considerably less formal education and more often held union membership than did the fathers of the Grace Line trainees.

The wide disparity between the ILGWU trainees' measured altruistic feelings and their

reported significant lack of success in actually dealing with other people is striking. The Grace Line trainees, according to middle-class standards, seem to be a more emotionally sound group, and reflect more clearly the stress placed on social maturity by the union and management program officials. The ILGWU *Ss* appear to reveal the strong social feeling sought by the union. They also seem to be aggressive, assertive, and impulsive, and perhaps, in meeting the entrance requirements, bring together such personality traits as would suggest, in an overall view, a lack of emotional maturity.

At the present time, no data pertaining to the *Ss*' relationships with each other, with their supervisors and with those with whom they must deal in the carrying out of their job duties is available. Accordingly, there are no external criteria immediately available against which the results of this study can be compared and, furthermore, any tendencies to answer questions according to a union or management stereotype, rather than to reflect their own true feelings, cannot be assessed.

Discussion

The literature has revealed a number of studies (Roe: 1950, 1952; Rieger, 1949; Daniels & Hunter, 1949) whose authors concluded that occupational personalities do exist, but who noted that caution must be exercised in putting their results to practical use. Individual differences within the groups investigated were often so great that though the mean trait scores significantly differed, the results had little operational utility.

Within-group differences, tending to minimize the practical significance of between-group differences, were frequently observed in this study. Though the majority of the hypotheses investigated produced statistically significant intergroup differences, the confidence with which the results could be put to use was somewhat impaired. Individual scores were frequently such that they could have placed *Ss* in either of the groups studied.

In describing the general kind of person who chooses to enter trade union leadership through the ILGWU training institute, or the field of business management through the Grace Line training program, the results should be valuable. The conclusions permit the construction of generalized typical ILGWU and Grace Line trainees and can provide some basis for individual counseling and for improved selection procedures.

The findings, however, though obviously of particular interest to the executives of the ILGWU and Grace Line, provide other more general implications.

It is apparent, in studies of the relationship between personality and occupational entry, that the efforts preceding and upon which instrument selection is based, are of uppermost importance. Though ideally the surveyed groups should first be studied by carefully controlled experimental methods, it has been shown, where the opportunity for lengthy pre-study investigations is not possible, that an effective test battery can be selected by first studying the subject group by means of careful, though not necessarily experimental methods, in which an emphasis is placed on teasing out the personality demands of the job. In the present investigation, instrument selection and the formulation of hypotheses followed discussions with executives of the two organizations studied, in which ideal types, based in part on generally desired characteristics and on recalled incidents "which distinguished the men from the boys," were constructed: the latter technique being, in some way, similar to the "critical incident technique" reported by Flanagan (1949).

As an aid to understanding better presently revealed personality traits, "relevant" early life experiences were also explored. Biographical inventory trait categories, referring to specific past behavior and family structure, were hypothesized in order to expose such antecedent experiences as may have significantly contributed to presently revealed personality traits and vocational choice. Though a psychoanalytic approach was employed in the postulation of particular life experiences, it seems obvious that regardless of the personality theory used present-day behavior can be more meaningfully understood when viewed within a developmental framework.

In most of the traits studied, the picture of the two groups revealed was surprisingly close to the ideal types sought by the particular union and management program administra-

tors. A comparison of the general characteristics of the Grace Line *S* with prevailing executive stereotypes and with the management trainees of another major industrial organization suggests that it is likely that they are in many ways representative of the occupation. It is probable, however, that the typical ILGWU trainee is not representative of the typical American trade union leader. The often reported increase in employment of well paid specialists by major American trade unions and data available in published biographies of leaders of other unions (Minton & Stuart [1937]; Wechsler [1944]; Madison [1950]) suggests that the avowed "dedication to trade union philosophies" and the acceptance by the ILGWU trainees of the fact that their earnings will never be high are just two of the many areas in which they may be observed to be atypical.

An empirical attempt to construct an occupational personality picture of the "trade union leader" would entail a study of considerably greater size and scope than the present one.

Significant is the consistency shown between *Ss*' early life experiences and present-day behavior, as suggested by the *Study of Values* and the *California Personality Inventory*. The role of societal pressures usually considered peculiar to certain social classes is also clearly demonstrated. Adding support to the literature suggesting a direct relationship between social stratification and personality, the management group, which on the *Sims Occupational Rating Scale* identified with a significantly higher social class, described early life experiences in which the conformity, child rearing techniques, and the stress of interpersonal skills experienced were considerably different from those encountered by the ILGWU students.

Though scoring significantly higher than the union group in areas of present-day aggression, the management students, consistent with their suggested early training experience, appeared to have channelized their drives into more socially acceptable areas. Instances of overt conflict with authority are less frequently seen, and the need to do things that are practical is strong.

Though both of the groups investigated were accepted for training only after being "carefully screened," within each of their training groups a few individuals were disclosed who seemed quite unlike their colleagues. Whether the personality characteristics of these variants will strongly influence their expected success is of obvious future interest, particularly as the homogeneity of each group is strong and the apparent differences between individuals and the groups norms are, in some instances, striking.

The study has added support to some general vocational theory and posed additional areas for further investigation. It has been demonstrated that significant personality differences exist between the occupational groups studied and that behavioral data, significantly differentiating two occupations and closely related to present-day personality and values, can be effectively elicited through custom-built biographical inventories.

Summary

In an attempt to study further the relationship of personality to occupational entry, the present investigation undertook to determine whether individuals who were considered worthy material for training for leadership positions by the ILGWU and the Grace Line, Inc., were different from each other.

Basing the selection of test instruments and hypotheses on job analyses and ideal types, which stressed personality demands of the job, 40 trainees in the formal training programs of each organization were studied. In order to investigate what were considered to be the most important personality traits, the *Sims SCI Occupational Rating Scale*, the Allport-Vernon-Lindzey *Study of Values*, the *California Psychological Inventory*, and a custom-built, trait-centered *Biographical Inventory* were employed and 15 hypotheses were formulated.

Of the 13 hypotheses which proposed differences in specified directions, seven were significant at less than the .01 level and one at less than the .05 level. Two hypotheses which did not specify intergroup differences, but proposed that the groups would score

high when compared with other occupational groups, were also supported.

Generally, the groups appeared to be composed of significantly different trainees. Individual differences within groups were, however, often considerable. In addition to supporting some general vocational theory, the results point to the practicability in occupational studies, when the opportunity for experimental prestudy is not possible, of preceding test selection with personality-factor job analyses, and to the discriminating ability of custom-built, trait-centered biographical inventories.

REFERENCES

DANIELS, E., & HUNTER, W. A. MMPI personality patterns for various occupations. *J. appl. Psychol.*, 1949, **33**, 559–569.

FLANAGAN, J. C. A new approach to evaluating personnel. *Personnel*, 1949, **26**, 35–42.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists, 1957.

MADISON, C. A. *American labor leaders*. New York: Harper, 1950.

MINTON, B., & STUART, J. *Men who lead labor*. New York: Modern Age, 1937.

PINTNER, R. A. A comparison of interests, abilities and attitudes. *J. abnorm. soc. Psychol.*, 1933, **4**, 351–357.

RIEGER, A. F. The Rorschach test and occupational personalities. *J. appl. Psychol.*, 1949, **33**, 572–579.

ROE, ANNE. Analysis of group Rorschachs of physical scientists. *J. proj. Tech.*, 1950, **14**, 385–398.

ROE, ANNE. Analysis of group Rorschachs of psychologists and anthropologists. *J. proj. Tech.*, 1952, **16**, 212–224.

SEGAL, S. J. The role of personality factors in vocational choice. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

WECHSLER, J. A. *Labor baron—a portrait of John L. Lewis*. New York: Morrow, 1944.

(Received May 18, 1959)

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:
UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:
J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

Timely

McGRAW-HILL

Books

PRINCIPLES OF COMPARATIVE PSYCHOLOGY

By ROLLAND H. WATERS, *University of Florida;* DOROTHY A. RETHLINGSHAFER, *University of Florida;* and WILLARD E. CALDWELL, *The George Washington University.* McGraw-Hill Series in Psychology. 453 pages, $7.95

This book represents the combined efforts of a group of distinguished psychologists working in the field of animal study. The purpose is to present a broad coverage of the field of animal behavior—to acquaint the reader with the methods and results of comparative studies and the principles to be drawn from them. Emphasis is placed upon current experimental and theoretical literature, without neglecting the standard material to be expected in such a survey.

QUANTITATIVE METHODS IN PSYCHOLOGY

By DON LEWIS, *State University of Iowa.* McGraw-Hill Series in Psychology. 558 pages, $9.50

A text for the increasingly important "Quantitative Methods" or "Quantitative Analysis" psychology courses. It is concerned with the application of nonstatistical, mathematical and graphical techniques to the clarification and interpretation of experimental data. The aim of the book is to provide a comprehensive survey of the quantitative and statistical procedures basic to the utilization of mathematical functions in describing empirical relationships and in developing theoretical schema for behavior.

EXPERIMENTS IN HEARING

By GEORG VON BEKESY, *Harvard University.* McGraw-Hill Series in Psychology. 744 pages, $25.00

An eminent international psychologist and physicist makes available to physiologists and psychologists the results of his unique ground-breaking experiments in the field of hearing. Over the years, von Bekesy has created his own experiments and has established physiological and neurological facts concerning hearing which have advanced this field considerably. The book is based on most of the papers published by the author during the last 30 years.

STATISTICAL ANALYSIS IN PSYCHOLOGY AND EDUCATION

By GEORGE A. FERGUSON, *McGill University.* McGraw-Hill Series in Psychology. 354 pages, $7.00

A textbook introducing the ideas and applications of statistics for students of the behavioral sciences—education, psychology, psychiatry, and sociology. Emphasis is placed on the analysis and interpretation of data resulting from the conduct of experiments with human and animal subjects. The book provides the technology necessary for the statistical treatment of most sets of experimental data encountered in practice.

Send for Copies on Approval

McGraw-Hill
BOOK COMPANY, INC.
330 West 42nd Street New York 36, N. Y.

PSYCHOLOGICAL ABSTRACTS

The basic reference journal in psychology

- Provides the key to psychological literature
- Noncritical abstracts of books and articles in the field of psychology and related subjects
- Abstracting and indexing an invaluable bibliographical aid
- The guide to what is going on in your science
- Wide foreign as well as U. S. coverage

NOW is the time to subscribe

Subscription	Single issue
$16.00 (foreign $16.50)	$3.00

Direct new subscriptions and renewals to:

American Psychological Association
1333 Sixteenth Street, N.W.
Washington 6, D. C.